Morphology in Auditory Lexical Processing

Sensitivity to Fine Phonetic Detail and Insensitivity to Suffix Reduction

ISBN: 90-76203-17-2
Printed and bound by Ponsen & Looijen bv, Wageningen

Morphology in Auditory Lexical Processing

Sensitivity to Fine Phonetic Detail and Insensitivity to Suffix Reduction

een wetenschappelijke proeve
op het gebied van de Letteren

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen,
op gezag van de Rector Magnificus Prof. dr. C. W. P. M. Blom,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op woensdag 1 september 2004
des namiddags om 3.30 uur precies

door

Rachèl Jenny Judith Karin Kemps
geboren op 23 augustus 1976 te Heerlen

Promotor:	Prof. dr. R. Schreuder
Co-promotores:	Dr. R. H. Baayen
	Dr. M. T. C. Ernestus (Max Planck Instituut voor Psycholinguïstiek, Nijmegen)
Manuscriptcommissie:	Prof. dr. C. Gussenhoven
	Prof. dr. L. Pols (Universiteit van Amsterdam)
	Dr. J. McQueen (Max Planck Instituut voor Psycholinguïstiek, Nijmegen)

Het onderzoek beschreven in deze dissertatie werd ondersteund door een PIONIER-subsidie van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), het Max Planck Instituut voor Psycholinguïstiek en de Faculteit der Letteren van de Radboud Universiteit Nijmegen.

Voor mijn ouders

Acknowledgements

I would like to express my gratitude to all those who have contributed to the incredible intellectual and personal enrichment that my dissertation years have been for me.

First of all, I would like to thank my promotor Rob Schreuder and my co-promotores Harald Baayen and Mirjam Ernestus for their invaluable guidance. Thank you for the opportunity to work with you and to learn from you. Thank you for motivating me and allowing me to attend an incredibly large number of workshops and conferences all over the world. Thank you for your unconditional support and confidence.

I would also like to acknowledge Lee Wurm, for his interest in my work and for giving me the opportunity to extend my research to English. This pleasant collaboration has resulted in a published paper and in Chapter 3 of this dissertation.

Furthermore, I owe many thanks to my fellow Ph.D students within the PIONIER project. Andrea Krott, Nivja de Jong, and Fermín Moscoso del Prado Martín, thank you for your friendship and support.

For their practical support, I am very indebted to all the research assistents that joined the PIONIER project over the last couple of years, to Lanneke van Dreumel, and to the technical and administrative staff of the Max Planck Institute.

I would like to acknowledge the members of the reading committee, Carlos Gussenhoven, James McQueen, and Louis Pols (in alphabetical order) for their critical and extremely helpful comments on the manuscript.

Furthermore, I would like to thank the members of the Comprehension Group of the Max Planck Institute for giving me the opportunity to attend their biweekly group meetings, and to benefit and learn from their knowledge and expertise.

Let me also acknowledge Gary Libben, for allowing me all the time and freedom that I needed to work on the final stages of my dissertation project.

I am very indebted to the Dutch Research Council (NWO), the Max Planck Institute for Psycholinguistics, and the Faculty of Arts of the University of Nijmegen for their financial support of my project.

Tenslotte, Martijn en Mirjam, dank jullie wel voor jullie steun in de afgelopen jaren en voor jullie hulp met de allerlaatste loodjes. Betere paranimfen had ik mij niet kunnen wensen. Elke, Elske, Evelien, Ingrid, Jules en Renske, bedankt voor jullie vriendschap. Papa, mama en Sven, dankjewel voor jullie vertrouwen in mij.

Edmonton, May 2004.

Contents

Introduction

CHAPTER 1

A purely segmental representation of the speech signal does not do justice to the extraordinary sensitivity of the speech-perception system on the one hand, and its enormous flexibility on the other hand. Many studies have shown that listeners are sensitive to the subsegmental acoustic details in the speech signal, and that these subsegmental cues can guide lexical activation (e.g., Davis, Marslen-Wilson, & Gaskell, 2002; Spinelli, McQueen, & Cutler, 2003; Salverda, Dahan, & McQueen, 2003). At the same time, speech perception does not seem to be hampered by the extreme reductions that typically occur in casual speech, even though these reductions sometimes involve the deletion of several segments or even syllables (Ernestus, 2000). This apparent contradiction poses questions on how words are represented and accessed in the mental lexicon. This thesis addresses some of these questions for the domain of morphological processing.

A great deal of previous experimental work on listeners' sensitivity to subsegmental information in the speech signal has focused on the role that subsegmental cues play in the competition between lexical candidates that show segmental overlap but that are morphologically unrelated. Davis et al. (2002) presented sentences that ended in either short words or long words that had these short words embedded at their onsets (e.g., *cap* or *captain*). They showed, in a gating task as well as in a cross-modal priming task, that the acoustic (durational and intonational) differences between short words and segmentally identical syllables embedded in longer words assisted the perceptual system in discriminating the short words from the start of longer words. Salverda et al. (2003) recorded participants' eye movements while they listened to Dutch sentences including a word with an onset-embedded word (e.g., *hamster* containing *ham*). There were more fixations to a picture representing the embedded word (*ham*) when the first syllable of the target word (*hamster*) had been replaced by a recording of the embedded word than when it came from a different recording of the target word. Subtle acoustic information in the speech signal appeared to lead the word-recognition system to favor the correct interpretation of lexically ambiguous spoken input.

This thesis focuses on the role of subsegmental cues in the processing of words that show segmental overlap *and* that are morphologically related. In Chapter 2, inflectionally related words are studied. Baayen, McQueen, Dijkstra, and Schreuder (2003) showed surface frequency effects in the auditory modality for nominal and verbal regular inflections in Dutch, suggesting the existence of full form representations for regularly inflected forms in the mental lexicon [1]. Most current models of spoken word recognition incorporate some form of lexical competition, such as the revised Cohort model (Marslen-Wilson, 1990; Marslen-Wilson, Moss, & Van Halen, 1996), TRACE (McClelland & Elman, 1986), and Shortlist (Norris, 1994). According to such lexical competition models, stored regularly inflected forms would be cohort competitors of their corresponding uninflected forms in languages in which the uninflected form is onset-embedded in the inflected form (e.g., Dutch, German, English): The uninflected form and the inflected form would keep on competing for recognition until after the offset of the stem. Inflectional embeddings are highly frequent: token-wise approximately equally frequent as morphologically unrelated embeddings, and type-wise approximately ten times as frequent as morphologically unrelated embeddings. For example, the Dutch stem *werk* ('work') is inflectionally embedded in the following forms: *werken, werkend, werkende, werkje, werkjes, werkt, werkte, werkten*. Because of the highly frequent occurrence of inflectional embeddings, the word-recognition system would benefit considerably from the presence and the functionality of acoustic cues that would distinguish the segmentally ambiguous portions of uninflected and inflected forms [2]. Chapter 2 investigates whether durational and intonational differences are present between Dutch

[1] Here and in the following, the term 'lexical representation' is used to refer to the phonological form representation of a word that is stored in the mental lexicon, and that, for recognition to occur, needs to be activated by the acoustic signal.

[2] Norris (1994) proposes another solution for the disadvantage that words with large word-initial cohorts suffer from: the 'cohort marker'. Norris discusses the 'cohort marker' idea for cohorts containing a large number of prefixed words, but the same idea could apply to cohorts containing stems and their inflectional variants: If the lexical search produces a large number of words that start with the same string of phonemes, this group of candidates can be replaced by a single entry in the candidate set, the cohort marker. This cohort marker competes for recognition just like any single word candidate. When new input fits one member of the cohort better than others, this word inherits the activation level of the cohort marker. The cohort marker provides an algorithmic solution for obtaining a relatively small shortlist that is underspecified as to when candidates in the initially merged cohort enter into the competition process.

singular nouns and the stems of their regularly formed plurals. The nouns under investigation are of the type *boek* - *boeken* ([buk] - [bukə(n)]; ‘book’ - ‘books’), that is, monosyllabic nouns that take the plural suffix *-en* [ə(n)]. In a number decision task, the combined and the independent effects of durational and intonational information in the speech signal on the processing of singular and plural forms are studied. Furthermore, by means of a lexical decision task, we investigate whether the prosodic differences between singulars and plurals are also picked up when the number of the noun is not relevant for the task at hand, and whether item-specific prosodic information might affect lexical processing.

Chapter 3 investigates whether the findings in Chapter 2 extend to different types of morphologically complex forms, and to another language, English. The morphologically complex forms under investigation in Chapter 3 are comparatives (inflection, e.g., Dutch: *vet* - *vetter*, [vɛt] - [vɛtəʀ]; English: *fat* - *fatter*, [fæt] - [fætɚ]) and agent nouns (derivation, e.g., Dutch: *werk* - *werker*, [wɛʀk] - [wɛʀkəʀ]; English: *work* - *worker*, [wɝk] - [wɝkɚ]). Studying the role of prosodic cues in the processing of comparatives and agent nouns in both Dutch and English enables us to determine whether the effects observed in the processing of singulars and plurals in Dutch (Chapter 2) are specific to plural formation in Dutch, or whether they generalize to a different type of inflection, to derivation, and to a different language. Dutch and English differ in morphological richness, in particular in the number of continuation forms that are possible given a certain monomorphemic stem. In English, a monomorphemic stem is less often followed by one or more unstressed suffixes than in Dutch. We investigate whether English listeners are, as a consequence, less sensitive than Dutch listeners to prosodic cues in the stem that signal whether or not the stem will be followed by unstressed syllables. Chapter 3 furthermore investigates the effects on processing of two covariates that are word-specific indications of the prevalence of possible continuation forms: Syllable Ratio and Cohort Entropy. Syllable Ratio is a lemma-based measure of the likelihood that a given stem will be onset-embedded in inflectional or derivational variants of that stem, consisting of the stem plus one or more unstressed syllables. Cohort Entropy is a cohort-based likelihood measure that a given stem will be embedded in a longer word that is or is not morphologically related to that stem. This approach enables us to determine whether there might be a word-specific component to an effect of the prosodic details in the speech signal.

Chapter 4 studies the role of subsegmental cues in the processing of Dutch sentences that are segmentally ambiguous, such as *Ik spreek de kerel soms* ('I sometimes talk to the guy', with the singular noun *kerel*) and *Ik spreek de kerels soms* ('I sometimes talk to the guys', with the plural noun *kerels*). Due to degemination of the cluster of [s]-es, the sequences *kerel soms* and *kerels soms* are segmentally identical: [keʀəlsɔms]. The ambiguous sentences studied in Chapter 4 contain monomorphemic nouns that take the plural suffix *-s* [s] (e.g., *kerel - kerels*, 'guy' - 'guys'), followed by words that have [s] as the onset. A first question addressed in Chapter 4 is whether nouns that take the plural suffix *-s* show durational differences between the singular and the stem of the plural similar to those observed for nouns that take the plural suffix *-en*, in ambiguous sentences as well as in non-ambiguous sentences (as opposed to in isolation). A second question is whether there might be other subsegmental differences that can reduce the ambiguity between *kerel soms* and *kerels soms*. Possibly, the degemination is incomplete: The contrast in the underlying phonological representations between one [s] and a cluster of two [s]-es might surface in the phonetic form as a difference in the duration of the [s] between *kerel soms* and *kerels soms* (see also Martens & Quené, 1994). Using a number decision task, we investigate which subsegmental cues — if any — listeners use to resolve the ambiguity between *kerel soms* and *kerels soms*.

Chapters 2 through 4 provide evidence for listeners' surprising sensitivity to the subtle details in the acoustic signal, and for a word-specific compononent to this sensitivity. Mismatching subsegmental information in the speech signal slows down processing. How then is it possible that comprehension is seemingly not hampered at all by the drastic reductions that typically occur in casual speech, when complete segments or syllables may be missing? This question is addressed in Chapter 5. Highly frequent words in particular may be drastically reduced. For example, the Dutch word *eigenlijk* ('actually') may be realized as [ɛiχələk], which is the canonical realization [3], but also as [ɛiχlək], [ɛiχlk], [ɛiχk], or [ɛik] (Ernestus, 2000). In the most extreme case [ɛik], the suffix *-e(n)lijk* [ələk] is reduced to a single [k]. Ernestus, Baayen, and Schreuder (2002) have shown that Dutch listeners recognize highly reduced word forms only when such forms are presented in a context of several words. Listeners do not recognize such forms when they are presented in isolation. Apparently, not all possible phonetic variants of a word are represented in the men-

[3] Here and in the following, the term 'canonical' is used to refer to the full, unreduced phonological realization/representation of a word.

tal lexicon or, if they are, they are not equally accessible. Furthermore, Ernestus et al.'s (2002) findings suggest that hearing reduced forms in context induces a process of reconstruction. Chapter 5 investigates, using a phoneme-monitoring task, whether reconstruction does indeed take place and — if so — what the precise nature of this reconstruction process might be.

References

Baayen, R. H., McQueen, J., Dijkstra, T., and Schreuder, R.: 2003, Frequency effects in regular inflectional morphology: Revisiting Dutch plurals, *in* R. H. Baayen and R. Schreuder (eds), *Morphological structure in language processing*, Mouton de Gruyter, Berlin, pp. 355–390.

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G.: 2002, Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition, *Journal of Experimental Psychology: Human Perception and Performance* **28**, 218–244.

Ernestus, M.: 2000, *Voice assimilation and segment reduction in casual Dutch, a corpus-based study of the phonology-phonetics interface*, LOT, Utrecht.

Ernestus, M., Baayen, R. H., and Schreuder, R.: 2002, The recognition of reduced word forms, *Brain and Language* **81**, 162–173.

Marslen-Wilson, W. D.: 1990, Activation, competition, and frequency in lexical access, *in* G. T. M. Altmann (ed), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, The MIT Press, Cambridge, Mass., pp. 148–172.

Marslen-Wilson, W. D., Moss, H. E., and van Halen, S.: 1996, Perceptual distance and competition in lexical access, *Journal of Experimental Psychology: Human Perception and Performance* **22**, 1376–1392.

Martens, L. and Quené, H.: 1994, Degemination of Dutch fricatives in three different speech rates, *in* R. Bok-Bennema and C. Cremers (eds), *Linguistics in the Netherlands 1994*, John Benjamins Publishing Company, Amsterdam.

McClelland, J. L. and Elman, J. L.: 1986, The TRACE model of speech perception, *Cognitive Psychology* **18**, 1–86.

Norris, D. G.: 1994, Shortlist: A connectionist model of continuous speech recognition, *Cognition* **52**, 189–234.

Salverda, A., Dahan, D., and McQueen, J. (2003), The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension, *Cognition* **90**, 51–89.

Spinelli, E., McQueen, J., and Cutler, A.: 2003, Processing resyllabified words in French, *Journal of Memory and Language* **48**, 233–254.

Prosodic cues for morphological complexity: The case of Dutch plural nouns

CHAPTER 2

This chapter has been tentatively accepted for publication as Rachèl J. J. K. Kemps, Mirjam Ernestus, Robert Schreuder, and R. Harald Baayen: Prosodic cues for morphological complexity: The case of Dutch plural nouns, *Memory and Cognition*.

Abstract

It has recently been shown that listeners use systematic differences in vowel length and intonation to resolve ambiguities between onset-matched simple words (Davis, Marslen-Wilson, & Gaskell, 2002; Salverda, Dahan, & McQueen, 2003). The present study shows that listeners also use prosodic information in the speech signal to optimize morphological processing. The precise acoustic realization of the stem provides crucial information to the listener about the morphological context in which the stem appears, and attenuates the competition between stored inflectional variants. We argue that listeners are able to make use of prosodic information, even though the speech signal is highly variable within and between speakers, by virtue of the relative invariance of the duration of the onset. This provides listeners with a baseline against which the durational cues in vowel and coda can be evaluated. Furthermore, our experiments provide evidence for item-specific prosodic effects.

Introduction

Several studies in the visual modality have shown surface frequency effects in the comprehension of fully regular inflections, thus providing evidence for storage of the inflected form as a whole at some level of representation. These effects have been shown for both nouns and verbs, and in several languages. For regularly inflected verbs, evidence for full form storage has been found for Dutch (Baayen, Schreuder, De Jong, & Krott, 2002; Schreuder, De Jong, Krott, & Baayen, 1999) and for English (Alegre & Gordon, 1999). For regularly inflected nouns, evidence for full form storage has been found for Dutch (Baayen, Dijkstra, & Schreuder, 1997), Finnish (Bertram, Laine, Baayen, Schreuder, & Hyönä , 1999), English (Sereno & Jongman, 1997; Alegre & Gordon, 1999), and Italian (Baayen, Burani, & Schreuder, 1997).

Recently, experiments in the *auditory* modality have also shown effects of full form frequency for both nominal and verbal regular inflections in Dutch, suggesting the existence of full form representations of regularly inflected forms in the auditory modality as well (Baayen, McQueen, Dijkstra, & Schreuder, 2003). This finding is surprising, given that most current models of spoken-word recognition incorporate some form of lexical competition, such as the revised Cohort model (Marslen-Wilson, 1990; Marslen-Wilson, Moss, & Van Halen, 1996), TRACE (McClelland & Elman, 1986), and Shortlist (Norris, 1994). In these models, stored regularly inflected forms would be cohort competitors of their corresponding uninflected forms: In many languages (e.g., Dutch, German, English), the uninflected form is onset-embedded in the longer inflected form and thus, at the phonemic level, the signal is ambiguous until the offset of the last phoneme of the stem (e.g., uninflected (singular) form: *book*; inflected (plural) form: *books*). In other words, the two candidates will keep on competing for recognition (i.e., in some models, inhibiting one another) until after offset of the uninflected form. Storage of regularly inflected forms creates a recognition problem in the domain of inflection, similar to the recognition problem that exists outside the domain of inflection — as for example in the perception of onset-embedded words that have longer morphologically unrelated competitors, such as *ham* in *hamster*.

Using the frequency counts in the CELEX lexical database (Baayen, Piepenbrock, & Van Rijn, 1993), we estimated how often both types of embedding (inflectional embedding versus morphologically unrelated embedding) occur in Dutch. We selected all 5129 monomorphemic lemmas that had a lemma frequency greater than zero. Subsequently, we selected all phonological word forms (uninflected and inflected forms) that corresponded to these lemmas. When we encountered a phonological form that contained an uninflected form at its onset, and that shared its stress pattern, we determined whether the stem of that form was the uninflected form. If so, we counted the phonological form as an inflectional continuation (e.g., [buk] - [bukə(n)], 'book' - 'books'). If the stem was not shared, we counted the phonological form as a morphologically unrelated continuation (e.g., [hɑm] - [hɑmstər], 'ham' - 'hamster'). This procedure resulted in the following counts: 2,188,144 tokens (307 types) were morphologically unrelated continuation forms, and 2,243,990 tokens (3015 types) were inflectional continuations [1]. These numbers show that inflectional embedding is a highly frequently occurring phenomenon: token-wise approximately equally frequent as morphologically unrelated embedding, and type-wise approximately ten times as frequent as morphologically unrelated embedding. The word-recognition system would therefore benefit considerably from the presence and the functionality of acoustic cues that would distinguish the segmentally ambiguous portions of uninflected and inflected forms.

In fact, evidence is accumulating that subtle subsegmental acoustic cues can reduce the ambiguity between onset-embedded words and their longer competitors, thus assisting the perceptual system in distinguishing them before the point in the acoustic signal at which disambiguating phonemic information comes in. Salverda, Dahan, and McQueen (2003) recorded participants' eye movements while they listened to Dutch sentences including a word with an onset-embedded word (e.g., *hamster* containing *ham*). The participants saw four pictures of objects on a computer screen and were instructed to use the computer mouse to move the picture of the object that was mentioned in the sentence. There were more fixations to a picture representing the embedded word (*ham*) when the first syllable of the target word (*hamster*) had been replaced by a recording of the embedded word than when it came from a different recording of the target word. Subtle acoustic information in the speech signal, namely, the duration of the embedded word relative to the duration of its corresponding syllable in the target word, appeared to lead the word-recognition system to favor the correct interpretation of lexically ambiguous

[1] These counts are based on the phonological transcriptions in the CELEX lexical database. These transcriptions explicitly code final devoicing of underlyingly voiced consonants. For example, the Dutch singular noun *hond* ('dog'), underlyingly ending in a voiced [d], is transcribed as [ɦɔnt], with a voiceless [t]. The plural *honden* ('dogs') is transcribed as [ɦɔndən], with a voiced [d]. The singular is thus not counted as embedded in the plural. Furthermore, these counts do not take differences in syllabic structure between longer words and their shorter embedded words into account. For example, both *boek-verkoper* ('book seller') and *boe-ken* ('books') were counted as continuation forms of *boek* ('book'), even though *boe-ken* does not share its syllabic structure with *boek*.

spoken input.

Experiments by Davis, Marslen-Wilson, and Gaskell (2002) similarly suggest that both durational and intonational differences assist the perceptual system in distinguishing short words from longer morphologically unrelated words beginning with these shorter words. In a gating task, participants were presented with sentence fragments. In one condition (long-word condition), the sentence fragments ended in a long carrier word of which the initial syllable formed an onset-embedded word (e.g., *captain* containing *cap*). In the other condition (short-word condition), the sentence fragments ended in the short word corresponding to the intitial syllable of the carrier word followed by a word with an onset that matched the continuation of the longer carrier word (e.g., *cap* t*ucked*, compare *cap*t*ain*). The first syllable in the short-word condition was significantly longer than the first syllable in the long-word condition, and there was a marginally significant difference in average fundamental frequency (average fundamental frequency was higher in the long-word condition than in the short-word condition). Significantly more short-word responses were made to gates from short-word stimuli than to gates from long-word stimuli, suggesting that listeners are influenced by the acoustic differences that exist between short and long word sequences in responding to the initial syllables of the target word. Similar results were obtained in a cross-modal priming task. The stimuli from the gating task were presented up to the offset of the first syllable of the target word (e.g., *cap* from either *cap* or *captain*) as auditory primes, and were followed by a visual target that was either the short word (*cap*) or the long word (*captain*). Greater facilitation occurred when prime syllables came from the same word as the target.

Outside the domain of inflection, listeners thus appear to be sensitive to durational and intonational differences between short words and longer lexical competitors. It is not self-evident that such prosodic differences are also sufficiently present to be functional for inflected words. Consider the Dutch singular and plural forms of ‘book’: *boek* [buk] and *boeken* [bukə(n)].

First, two phonetic processes exert their influence in parallel: a *shortening* process and a *lengthening* process. For Dutch, the shortening process has been described by Nooteboom (1972). In a stress-timed language like Dutch, the duration of a stressed vowel reduces as a function of the number of unstressed syllables that follow (see also: Lehiste, 1972, and Fowler, 1977, for English; Lindblom & Rapp, 1973, for Swedish). Therefore, the duration of the vowel in the first syllable in *hamster* is expected to be shorter than the duration of the same vowel in *ham*. The same holds for the vowel in the first syllable in *boeken* as compared to the same vowel in *boek*. However, since the second syllable in *boeken* is less complex than the second syllable in *hamster*, it is conceivable that the amount of shortening in words like *boeken* versus *boek* is smaller compared to the amount of shortening in words like *hamster* versus *ham*. The amount of shortening might not be enough to be functional for the listener.

Simultaneously, a prosodic lengthening process applies: The last syllable before a prosodic boundary (e.g., a prosodic word boundary or a phonological phrase boundary) is lengthened. Therefore, the form *ham* (which is followed by a word boundary) is expected to be longer than the first syllable in *hamster* (which is not followed by a word boundary). Cambier-Langeveld (2000) points out that when the rhyme of the last syllable consists of a schwa, as for example in words like *boeken* [bukə(n)], prosodic lengthening also applies to the penultimate syllable. In other words, in *hamster* only the last syllable is subject to prosodic lengthening, whereas in *boeken* both syllables are lengthened. Thus, it is likely that the difference between *boek* and the first syllable of *boeken* is smaller than the difference between *ham* and the first syllable of *hamster*.

Bearing these phonetic considerations in mind, it is not self-evident that durational modification of the first syllable occurs in inflected forms to the same extent as it does in words carrying onset-embedded morphologically unrelated words. The durational modification in inflected forms might not be sufficiently present to be functional.

Linguistic considerations lead to the same conclusion. Various linguists have argued that it is preferable for the phonological form of the stem to remain unaltered after affixation. For instance, Aronoff (1976) points out that affixes that leave their base words unchanged, i.e., that are phonologically transparent, are more productive than affixes that lead to phonological opacity (see also Dressler, Mayerthaler, Panagl, & Wurzel, 1987, for morpho-phonological processes in general). In Optimality Theory, this idea is implemented by means of alignment constraints between prosodic and morphological constituents (e.g., McCarthy & Prince, 1993). These linguistic considerations lead one to expect that it would be dysfunctional for the stem in isolation to differ from the stem followed by an inflectional ending.

Considered jointly, these phonetic and linguistic considerations show that it is not obvious that systematic subsegmental differences between inflected forms and their base words might exist and be functional for the listener.

On the other hand, the existence of functional prosodic differences in the domain of inflection would reduce the competition problem created by the storage of regular inflected forms in the auditory modality. Such differences would distinguish uninflected forms from their longer inflectional counterparts well before the offset of the uninflected form — their uniqueness point would then occur considerably earlier than suggested by their phonemic representation.

Interestingly, an indication that subsegmental differences may exist between uninflected forms and their longer inflectional counterparts is provided in a pilot study by Baayen et al. (2003). The singular and plural forms of five nouns were realized five times by four native speakers of Dutch. The mean duration of the singulars was longer (98 ms on average) than that of the stems embedded in the plurals.

The question arises whether such prosodic cues in the domain of inflection can be functional for the listener, given the enormous variability of speech within and across speakers. In the present paper, we address this question by means of an experimental study of regular plural nouns in Dutch. In Dutch, the regular plural form of many nouns consists of the noun stem and the plural suffix *-en* (usually realized as just a schwa; e.g., *boek* [buk] 'book' – *boeken* [bukə] 'books'). We studied both the combined and the independent effects of durational and intonational information in the speech signal on the processing of singular and plural forms, using a number decision task as well as a lexical decision task. We furthermore investigated whether item-specific prosodic information might affect lexical processing.

Experiment 2.1

The question addressed in Experiment 2.1, employing a number decision task, is whether listeners are sensitive to prosodic differences between singular forms and the stems of plural forms. If so, listeners are expected to be slowed down in their responses when there is a mismatch between the prosodic (durational and intonational) information in the acoustic signal of a word on the one hand, and the word's number as it is conveyed by the presence or absence of the plural suffix on the other hand. Moreover, the magnitude of the delay in response latencies is expected to covary with the degree of prosodic mismatch. We will test the covariance between degree of prosodic mismatch and magnitude of the delay in response latencies in a regression design. If listeners are not sensitive to prosodic differences between singular and plural forms, in other words, if listeners rely on segmental information only, mismatching prosodic information should not affect response latencies.

Method

Participants. Forty-six participants, mostly students at the University of Nijmegen, were paid to participate in the experiment. All were native speakers of Dutch.

Materials. From the CELEX lexical database (Baayen et al., 1993) we selected all Dutch monomorphemic nouns that met the following criteria: Their initial phoneme was not a vowel, their plural was formed by adding the suffix *-en* [ə(n)] to the stem, and they did not also function as verbal forms. Furthermore, they ended in an underlyingly voiceless plosive. In Dutch, the rule of final devoicing applies: Underlyingly voiced obstruents in syllable-final position are devoiced. The plural suffix *-en* [ə(n)] induces resyllabification of the stem-final obstruent as onset of the next syllable, and hence an underlyingly voiced stem-final obstruent will remain voiced (Booij, 1995). As a consequence, only stems ending in underlyingly voiceless obstruents phonemically have the same base in the singular as in the plural form. We therefore only selected nouns with stems ending in an underlyingly voiceless plosive, so that there is no change of the voicing characteristics of the plosive when the stems occur in isolation. Finally, the singular surface frequencies and plural surface frequencies of the nouns were larger than zero. (Singular surface frequency and plural surface frequency are token counts. Token counts in CELEX are based on a corpus of 42 million words of written Dutch.) From the resulting group of 135 nouns, we selected 48 experimental nouns that contained a simplex coda. These nouns are listed in Appendix A. Additionally, we randomly selected 48 filler nouns from the group of 133 Dutch monomorphemic nouns that met all the above criteria, except that these nouns could also function as verbal forms.

We excluded nouns containing a complex coda for the following reason. As mentioned above, for stress-timed languages, the vowel duration in a stressed syllable decreases as a function of the number of unstressed syllables that follow (Nooteboom, 1972, for Dutch; Lehiste, 1972, and Fowler, 1977, for English; Lindblom & Rapp, 1973, for Swedish). This effect of the number of following syllables is smaller with smaller vowel duration in the stressed syllable (Nooteboom, 1972; Lehiste, 1972). In other words, the smaller the vowel duration in the monosyllabic singular form, the smaller the difference that is to be expected between the vowel duration in the singular form and the vowel duration in the bisyllabic plural form. Since vowels have a smaller duration when they are followed by a complex coda than when they are followed by a single consonant (Waals, 1999, for Dutch; Munhall, Fowler, Hawkins, & Saltzman, 1992, for English), the difference between singular and plural forms is expected to be smaller for words ending in a complex coda than for words ending in a single consonant. We decided to exclude nouns with a complex coda, so that the durational difference to be expected between the vowel in the singular form and the vowel in the plural form was maximal.

Three reading lists were created: a list containing the singular forms of the experimental nouns, a list containing the plural forms of the experimental nouns, and a list containing the plural forms of the filler nouns. The order of the nouns within lists was randomized three times, resulting in 9 reading lists. In order to maximize durational differences between singular and plural forms, the noun forms were read in isolation. The lists were recorded in a soundproof recording booth by a native female speaker of Dutch, who was naive regarding the purpose of the experiment. The recordings were digitized at 16 kHz.

For each noun form, the best realization (of three) was selected and spliced out of its list using the PRAAT speech-editing software (Boersma & Weenink, 1996). Subsequently, out of the experimental noun forms we created two types of singular forms: 'normal' singular forms and 'constructed' singular forms. The normal singular form consisted of the singular form exactly as it was uttered by the speaker. The constructed singular form consisted of the stem of the plural form — in other words, it was the plural form with the plural suffix *-en* [ə(n)] spliced off. The point of splicing was located at the onset of the voicing of the schwa following the stem-final consonant. As a result, the normal singular form's prosodic information matched its number information (as conveyed by the absence of the plural suffix), whereas the constructed singular form's prosodic information mismatched its number information: Its prosodic characteristics signalled a plural form, whereas the absence of the plural suffix signalled a singular form. Total duration, vowel duration, closure duration, and release noise duration were measured for the two types of singular forms. Onset of the vowel was defined as onset of voicing if the preceding segment was voiceless, and as the end of the release noise if the vowel followed a fully voiced stop. In all other cases (i.e., if the preceding segment was [l, ʀ, m, n, v] or [ʋ]), onset of the vowel was defined as the point of change in the periodicity pattern in the waveform. The end of the vowel and beginning of closure was defined as the end of the second formant of the vowel. The end of closure was located at the onset of the sudden discontinuity in the waveform for the release noise. A paired *t*-test showed that on average the constructed singular forms had a significantly smaller total duration than the normal singular forms ($t(47) = 18.2, p < 0.0001$). The mean difference in total duration was 87 ms. The mean difference in vowel duration was

17 ms ($t(47) = 14.8, p < 0.0001$), the mean difference in closure duration was 26 ms ($t(47) = 10.9, p < 0.0001$), and the mean difference in release noise duration was 37 ms ($t(47) = 13.8, p < 0.0001$). An analysis of variance with total duration as the dependent variable showed that there was no significant interaction between type of singular form (normal versus constructed singular form) and type of vowel (phonologically and phonetically long versus short vowel): The difference in duration between normal and constructed singular forms was comparable for words with phonologically and phonetically long and short vowels ($F(1, 92) = 0.4, p = 0.52$). Table 2.1 lists the mean durations with their standard deviations for the two kinds of singulars.

Table 2.1: Experiment 2.1 – Mean durations (in ms) with *SD* for normal and constructed singular forms.

	Normal singular form		Constructed singular form		Duration difference
	Duration	SD Duration	Duration	SD Duration	
Whole form	388	73	301	73	87
Vowel	138	45	121	42	17
Closure	88	21	62	14	26
Release noise	76	24	39	15	37

Furthermore, we measured the average fundamental frequencies of the normal and the constructed singular forms. Recall that Davis et al. (2002) found that the average fundamental frequency was higher in the initial syllables of bisyllabic words than in monosyllabic words. We obtained a similar result: The constructed singular forms had a significantly higher average fundamental frequency than the normal singular forms ($t(47) = -2.0, p < 0.05$). The mean difference in average fundamental frequency was 7 Hz (185 Hz for the normal singular forms and 192 Hz for the constructed singular forms). Our explanation for this finding is that all (monosyllabic and bisyllabic) forms were pronounced with an intonational phrase final pitch accent H*L, which was aligned differently in monosyllabic than in bisyllabic words. In the case of the monosyllabic forms, both H and L were realized within one syllable. In the case of the bisyllabic forms, H was assigned to the first (stressed) syllable and L was assigned to the second syllable. Consequently, average fundamental frequency was higher in the first syllables of the bisyllabic forms than in the monosyllables.

The normal and constructed singular forms functioned as experimental target words. Filler words consisted of the plural filler nouns, exactly as they were uttered by the speaker.

Three experimental trial lists and their complements were created in such a way that each list contained all 48 filler items (plural forms), 24 normal singular forms, and 24 constructed singular forms. One list never contained both the normal and the constructed singular form of a single noun: If a given list contained the normal singular form of a noun, then the constructed singular form of that noun was contained in its complementary list. The order of presentation of the stimuli was pseudo-randomized within the three lists: No more than three singular forms of the same type occurred successively. Orders were identical in complementary lists. Participants were randomly assigned to experimental trial lists. Practice trials were presented prior to the actual experiment. The practice set consisted of 16 trials: 8 plural forms, 4 normal singular forms, and 4 constructed singular forms (all taken from a different recording of the complete experimental and filler sets). None of the nouns in the practice set was presented in the actual experiment.

Procedure. Participants were instructed to decide as quickly as possible whether the form they heard was a singular or a plural form. They responded by pressing one of two buttons on a button box. All experimental items required the response 'singular', whereas all filler items required the response 'plural' (assuming that decision on number is primarily based on the presence versus absence of a plural suffix). Each trial consisted of the presentation of a warning tone (377 Hz) for 500 ms, followed after an interval of 450 ms by the auditory stimulus. Stimuli were presented through Sennheiser headphones. Reaction times were measured from stimulus offset. Each new trial was initiated 2500 ms after offset of the previous stimulus. When a participant did not respond within 2000 ms post-offset, a time-out response was recorded. Prior to the actual experiment, the set of practice trials was presented, followed by a short pause. The total duration of the experimental session was approximately 10 minutes.

Results and discussion

No participants or items were excluded from the analyses, since they all showed error rates below 20%. In all analyses, only trials eliciting correct responses were included. The mean reaction times (measured from word offset, and calculated over the correct trials only) and the percentages of incorrect trials for the two kinds of singulars are summarized in Table 2.2.

If listeners are sensitive to prosodic differences between singular and plural forms, our dependent variable reaction time should covary with the degree of prosodic mismatch between normal and constructed singular forms. Simply finding a

Table 2.2: Experiment 2.1 – Mean response latencies (in ms) measured from word offset (calculated over correct trials only) with *SD* and percentages of incorrect trials for normal singular forms and constructed singular forms.

Type of singular form	RT	SD	% Incorrect
Normal	335	44	2.0
Constructed	444	36	1.4

delay in processing (109 ms; $t1(45) = -16.0, p < 0.0001$; $t2(47) = -15.4, p < 0.0001$) is not sufficient evidence for the occurrence of a prosodic mismatch effect, as this delay might as well be a consequence of the splicing manipulation that has been applied to the constructed singular forms. What needs to be shown is a correlation between the magnitude of the prosodic mismatch and the delay in processing.

We therefore applied a covariance analysis along the lines of Lorch and Myers (1990) to the reaction time data corresponding to the constructed singular forms. We operationalized the amount of prosodic mismatch as the difference between the duration of the constructed singular form and the duration of the corresponding normal singular form. As mismatch in intonational contour is not straightforwardly quantifiable — average fundamental frequency does not capture contour information — we did not include intonational mismatch in the numeric operationalization of prosodic mismatch. It is conceivable, however, that the amount of intonational mismatch codetermined reaction times to the constructed singular forms, and we will return to this issue below. We fitted a linear model to the data for each participant separately, with log reaction time as the dependent variable, and log singular surface frequency, duration of the form itself, and the durational difference score as predictors. Singular surface frequency was included as a predictor in order to ascertain that any observed effect of the durational difference score could not be a consequence of confounding durational differences with differences in frequencies between the items. T-tests on the coefficients of the participants for the predictors revealed that duration had a facilitatory effect (the longer the duration, the shorter the response latencies; $t(45) = -3.0, p < 0.01$), and durational difference had an inhibitory effect (the larger the durational mismatch, the longer the response latencies; $t(45) = -3.0, p < 0.01$).

Using a multi-level extension of the Lorch and Myers-technique (Pinheiro & Bates, 2000), we tested whether any effect of durational difference remained after partialling out the effects of the other predictors. This was indeed the case ($F(1, 1035) = 6.0, p < 0.05$), indicating that durational difference had an independent effect on the reaction times to the constructed singular forms [2].

Apparently, when listeners segmentally perceive a singular form, but prosodically (durationally) a plural form is signalled, their number decision is adversely affected. What then happens in the opposite situation? What happens when segmentally a plural form is presented, but prosodic cues in the stem signal a singular form? In Experiment 2.2 we investigated whether we may replicate this prosodic mismatch effect for plural forms. We created two types of plural forms: one form in which the prosodic (durational and intonational) cues *matched* the number of the form as it was conveyed by presence of the suffix, and one form in which the prosodic cues *mismatched* the number of the form as conveyed by the presence of the suffix.

Experiment 2.2

Method

Participants. Forty-three participants, mostly students at the University of Nijmegen, were paid to participate in the experiment. All were native speakers of Dutch. None of them had participated in Experiment 2.1.

Materials. The target items in this experiment were normal and constructed *plural* forms. Contrary to in Experiment 2.1, both types were now created by means of a splicing manipulation, which allows a factorial experimental design contrasting normal and constructed forms. The filler items were now singular forms.

We selected the same experimental and filler nouns as in Experiment 2.1. The singular forms of the experimental nouns, the plural forms of the experimental nouns, and the singular forms of the filler nouns were assigned to separate reading lists. The order of the nouns within lists was randomized three times, resulting in 9 reading lists. These lists were read by the same native female speaker as in Experiment 2.1. The lists were recorded in a soundproof recording booth. The recordings were digitized at 16 kHz.

[2]In the analyses reported here, log reaction times were predicted as measured from word offset. A model predicting log reaction times as measured from word *onset* yielded the following results: an inhibitory effect of duration (the longer the duration, the longer the response latencies: $t(45) = 20.0, p < 0.0001$), and an inhibitory effect of durational difference (the larger the durational mismatch, the longer the response latencies: $t(45) = -2.7, p < 0.05$). The effect of durational difference remained significant after partialling out the effects of the other predictors ($F(1, 1035) = 4.1, p < 0.05$).

Subsequently, we created the two types of plural forms: normal plural forms and constructed plural forms. Both types of plural forms were created using a splicing technique: The beginning of one speech token was combined with the ending of a different speech token. From both the singular and the plural form of a noun, we selected the portion of signal from the first phoneme up to and including the closure of the final plosive of the stem. In other words, we selected the stem without the release noise of the final plosive. From another realization of the plural form of the same noun, we selected the portion from the release noise of the final plosive of the stem up to and including the last phoneme. To create the normal plural form, we concatenated this latter portion to the initial portion originating from the plural from. To create the constructed plural form, we concatenated it to the initial portion originating from the singular form. As a result, the normal plural form consisted of two portions of signal both originating from plural forms, whereas the constructed plural form consisted of an initial portion originating from a singular form and a final portion originating from a plural form. This splicing manipulation is illustrated in Figure 2.1. Note that by applying this splicing procedure to both the normal and the constructed plural forms, we ensured that any observed difference in response latencies cannot be a consequence of a difference in splicing manipulation: A delay in processing for the constructed plural forms would constitute sufficient evidence for the occurrence of a prosodic mismatch effect.

Since the initial portion of the constructed plural form originated from a singular form, it was expected to contain durational and intonational cues that mismatched the number of the word as it was conveyed by the presence of the plural suffix. A paired *t*-test indeed showed a significant difference in total duration between the normal and the constructed plural form: The constructed plural form was longer (29 ms on average) than the normal plural form ($t(47) = 5.6, p < 0.0001$). The difference in vowel duration was 15 ms ($t(47) = 6.6, p < 0.0001$) and the difference in closure duration was 19 ms ($t(47) = 6.4, p < 0.0001$). Table 2.3 lists the mean total durations with their standard deviations for the two types of plural forms.

Table 2.3: Experiment 2.2 – Mean durations (in ms) with *SD* for normal and constructed plural forms.

	Normal plural form		Constructed plural form		Duration difference
	Duration	SD Duration	Duration	SD Duration	
Whole form	487	77	516	75	29
Vowel	117	45	132	42	15
Closure	65	17	84	23	19

In addition, intonational differences were present between the initial portions of

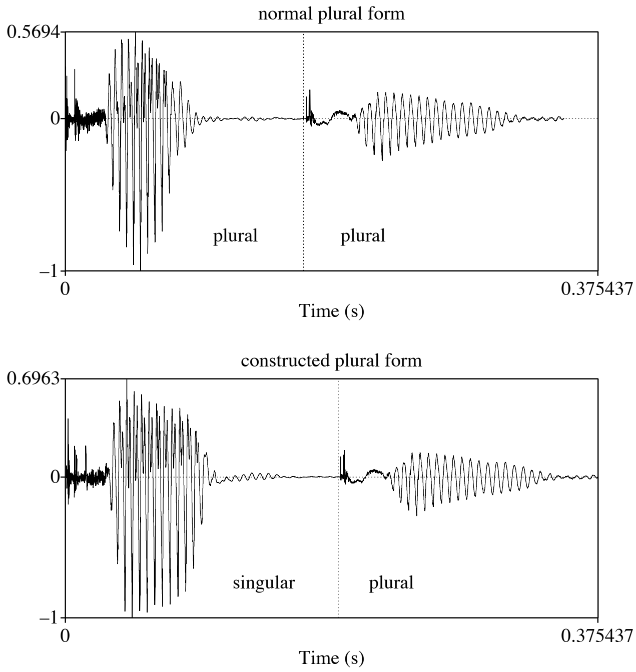

Figure 2.1: A normal plural form, consisting of two portions of signal originating from plural forms (upper panel), and a constructed plural form, consisting of an initial portion originating from a singular form and a final portion originating from a plural form (lower panel).

the normal and constructed forms: The average fundamental frequency of the initial portion of the constructed plural form was on average 11 Hz lower than the average fundamental frequency of the initial portion of the normal plural form ($t(47) = -10.9, p < 0.0001$; 190 Hz for the normal plural forms and 179 Hz for the constructed plural forms).

In the case of the constructed plural forms, this splicing procedure gave rise to artificial plosives that combined the closure of a singular realization with the release noise of a plural realization. Or, put differently, durational information contained in the original release noise of the singular realization was no longer present in the acoustic signal of the constructed plural form. Recall that we applied this splicing procedure in order to ensure that any observed difference in response latencies cannot be a consequence of a difference in splicing manipulation. But would it have been more natural, and more analogous to the creation of the constructed singular forms in the previous experiment, to simply concatenate the plural suffix to the singular stem when forming constructed plural forms? Actually, it turned out that this latter procedure gave rise to very unnaturally sounding stimuli. In fact, this by itself already exactly answers our research question: A plural form is not simply a singular form with a plural suffix concatenated to it, neither in production nor in perception. The stem in the plural form contains acoustic cues that distinguish it from the same stem in the singular form. In order to prevent participants from showing unnatural behavior as a result of the presence of very unnaturally sounding stimuli in the experiment, and in order to determine whether prosodic cues other than the nature of the release noise play a role in the processing of plurals, we opted for the present, somewhat more complicated cross-splicing procedure.

Three trial lists and their complements were created in the same manner as in Experiment 2.1: Each list contained all 48 filler items (singular forms), 24 normal plural forms, and 24 constructed plural forms. Participants were randomly assigned to experimental trial lists. Practice trials were presented prior to the experiment. The practice set consisted of 16 trials: 8 singular forms, 4 normal plural forms, and 4 constructed plural forms. None of the nouns in the practice set was presented in the actual experiment.

Procedure. The same experimental procedure was used as in Experiment 2.1, except that now all experimental items required the response ‘plural’ and all filler items required the response ‘singular’ (again assuming that number decision is primarily based on the presence versus absence of a plural suffix).

Results and discussion

All items and subjects were included in the analyses, since they all showed error rates below 20%. Table 2.4 lists the mean reaction times (calculated over the correct trials only) and the percentages of incorrect trials for the two types of plural forms.

Table 2.4: Experiment 2.2 – Mean response latencies (in ms) measured from word offset (calculated over correct trials only) with *SD* and percentages of incorrect trials for normal plural forms and constructed plural forms.

Type of plural form	RT	SD	% Incorrect
Normal	299	50	1.2
Constructed	323	53	1.0

Paired *t*-tests showed a significant difference in response latencies: Response latencies to the constructed plural forms were longer (24 ms on average) than to the normal plural forms ($t1(42) = -3.6, p < 0.001$; $t2(47) = -2.3, p < 0.05$). The physically longer items thus produced the longer reaction times. A simple processing explanation (i.e., longer signal to process, longer processing time), however, seems rather unlikely, since reaction times were measured from word offset. Furthermore, the covariance analysis in Experiment 2.1 shows that duration in fact has a facilitatory effect: Longer item durations were associated with shorter reaction times.

Instead, the prosodic mismatch effect originally observed for singular forms appears to have occurred for plural forms as well. Interestingly, the effect for the plural forms was considerably smaller than the effect for the singular forms (24 ms for the plurals in Experiment 2.2 as opposed to 109 ms for the singulars in Experiment 2.1). There are three possible explanations for this.

First, the magnitude of the prosodic mismatch was larger for the singulars in Experiment 2.1 than for the plurals in Experiment 2.2 ($t(47) = -9.1, p < 0.0001$). Whereas in Experiment 2.1 all durational information carried by the stem of the plural form was contained in the constructed singular form, in Experiment 2.2 durational information contained in the release noise of the final plosive of the singular was no longer present in the constructed plural form as a consequence of the splicing procedure. An explanation of the different delay magnitudes between experiments in terms of different mismatch magnitudes is supported by the fact that there was an inhibitory effect of durational difference in Experiment 2.1 ($F(1, 1035) = 6.0, p < 0.05$) as well as in Experiment 2.2 ($F(1, 999) = 26.1, p < 0.0001$).

Second, the nature of the expectancy violation in Experiment 2.1 was different from that in Experiment 2.2. In Experiment 2.1, presentation of the constructed singular form led the listener to expect a plural form on the basis of the durational (and possibly intonational) cues that were present in the signal, but then suddenly the auditory signal was broken off, leaving the listener with conflicting evidence. In Experiment 2.2, presentation of the constructed plural form initially led the subjects to expect a singular form, but then the auditory signal continued until the end of the plural suffix. Evidence pointing to the plural form thus kept accumulating after the stem, partly compensating for the subtle prosodic cues in the stem pointing to the singular form. It is possible that this difference in the nature of the violation of the expectancy was also reflected in the different magnitudes of the prosodic mismatch effect in response latencies.

Finally, it is possible that the difference in delay magnitudes between Experiment 2.1 and Experiment 2.2 is a result of the fact that, in Experiment 2.1, the manipulation of interest had been systematically confounded with the splicing manipulation. Thus, the delay observed for the constructed forms in Experiment 2.1 may have partly been the result of the splicing manipulation applied to these forms. There was no such splicing confound in Experiment 2.2. We cannot rule out this possibility based on our results, but we would like to stress here that the crucial finding in Experiment 2.1 was not the delay per se, but the relation between the magnitude of the durational differences and the response latencies. This relation shows that the delay observed in Experiment 2.1 cannot solely be attributed to the splicing manipulation.

The covariance analyses described under Experiment 2.1 and 2.2 showed that reaction times to the constructed singular forms in that experiment were at least partly determined by the magnitude of the durational mismatch between the normal and the constructed forms. As mentioned before, mismatch in intonational contour is not as easily quantifiable, and can therefore not similarly be included as a predictor in a linear model. We therefore investigated the individual contribution of intonational information to the prosodic mismatch effect in a separate experiment. In Experiment 2.3, again, normal and constructed singular forms were presented, but now these two types of singular forms only differed in intonational contour. If intonational cues contribute to the prosodic mismatch effect, we should observe longer response latencies to the forms with the mismatching intonational contour.

Experiment 2.3

Method

Participants. Forty-nine participants, mostly students at the University of Nijmegen, were paid to participate in the experiment. All were native speakers of Dutch. None of them had participated in Experiment 2.1 or 2.2.

Materials. The normal singular forms from Experiment 2.1 were used with no further manipulation. In addition, new constructed singular forms were created by taking the normal singular forms, and overlaying them with the intonational contours taken from the stems of the plural forms. This manipulation was carried out using the PSOLA (Pitch-Synchronous Overlap and Add) resynthesis method in the PRAAT speech-editing program (Boersma & Weenink, 1996). Figures 2.2, 2.3, and 2.4 illustrate the manipulation of the intonational contour. The intonational contour of the singular stem in the lower panel of Figure 2.2 was combined with the waveform of the plural stem in the upper panel of Figure 2.3, resulting in the singular stem with the plural intonational contour as displayed in Figure 2.4. As a result, the durations of the two types of singular forms were identical, but one type of singular form carried the intonational contour of the singular ('normal' singular form), whereas the other type of singular form carried the intonational contour of the plural ('constructed' singular form). The same filler words were used as in Experiment 2.1.

Three trial lists and their complements were created in the same manner as in the previous experiments: Each list contained all 48 filler items, 24 normal singular forms, and 24 constructed singular forms. Participants were randomly assigned to experimental trial lists. Practice trials were presented prior to the actual experiment. The practice set consisted of 16 trials: 8 plural forms, 4 normal singular forms, and 4 constructed singular forms. None of the nouns in the practice set was presented in the actual experiment.

Procedure. The same experimental procedure was followed as in the previous experiments.

Results and discussion

We included all items and participants in the analyses, since they all showed error rates below 20%. Table 2.5 lists the mean reaction times (calculated over the cor-

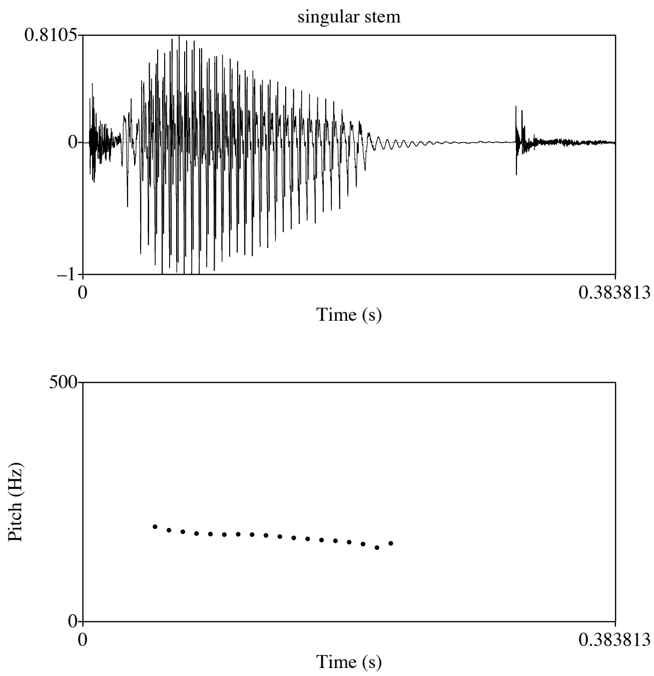

Figure 2.2: The waveform (upper panel) and the intonational contour (lower panel) of a singular stem.

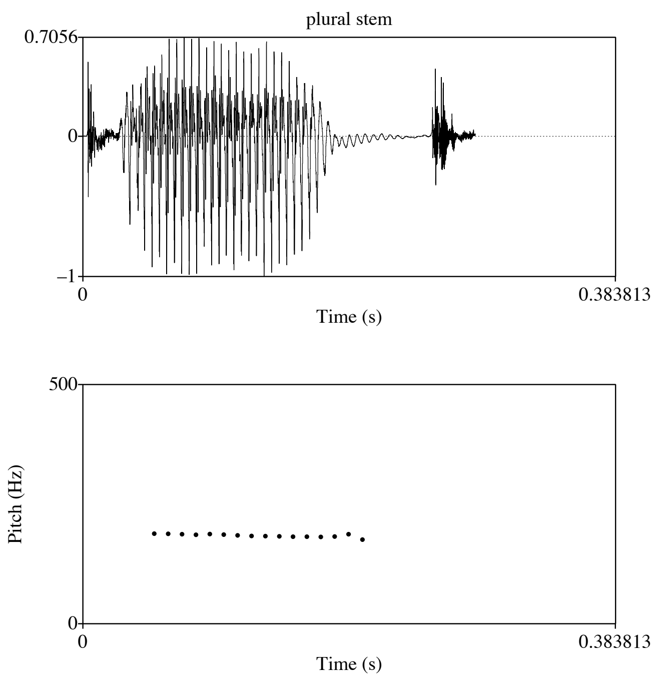

Figure 2.3: The waveform (upper panel) and the intonational contour (lower panel) of a plural stem.

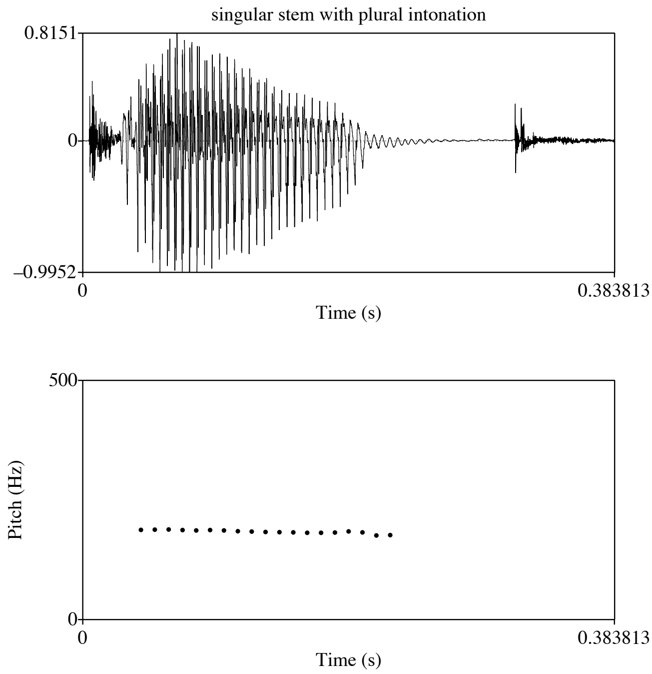

Figure 2.4: A singular stem with a plural intonational contour, resulting from combining the waveform of Figure 2.2 with the intonational contour of Figure 2.3.

rect trials only) and the percentages of incorrect trials for the two kinds of singular forms.

Table 2.5: Experiment 2.3 – Mean response latencies (in ms) measured from word offset (calculated over correct trials only) with *SD* and percentages of incorrect trials for normal singular forms and constructed singular forms when the two types of singular forms only differed in intonational contour.

Type of singular form	RT	SD	% Incorrect
Normal	333	51	1.7
Constructed	343	47	0.9

Participants responded on average 10 ms slower to the constructed singular forms than to the normal singular forms. In a paired *t*-test, this difference was significant by subjects ($t1(48) = -2.2, p < 0.05$), but not by items ($t2(47) = -1.5, p = 0.14$). As the item-analysis may be too conservative for the type of experimental design used in this study (Raaijmakers, Schrijnemakers, & Gremmen, 1999), we additionally ran a covariance analysis (Lorch & Myers, 1990), in which the factor type of singular form (normal singular form versus constructed singular form) and the covariate log singular surface frequency predicted log reaction times. This analysis revealed significant effects of both type of singular form ($t(48) = -2.8, p < 0.01$)

and of singular surface frequency ($t(48) = -4.7, p < 0.0001$).

These results suggest that when intonational information mismatches number information (conveyed by the presence/absence of the plural suffix), number decision is hindered. Both duration and intonation thus appear to serve as cues in perceptually distinguishing between singular and plural forms. The processing delay for stimuli with mismatching intonational contour was only 10 ms. Note, however, that the stimuli in our experiments were presented in isolation. The participants did not hear surrounding speech that could function as a frame of reference against which they could evaluate the fundamental frequency of the stimuli. It is conceivable that, when singulars and plurals are presented in their context, intonation serves as a considerably stronger cue than it did in this experiment. An alternative explanation for the relatively small effect of intonational mismatch on reaction times is that the intonational difference is peculiar to the context in which the words were produced — contrary to the durational difference, which is probably quite systematically present between singulars and plurals produced in any context. In a list context, each word will have an intonational phrase final contour. This contour will be aligned differently for monosyllables than for bisyllabic forms, leading to differences in average fundamental frequency in the first syllable. However, singulars and plurals do not typically occur in phrase final position, and will therefore not show differences in average fundamental frequency as systematic as the durational differences. If intonational differences are indeed less systematic than durational differences, it is not surprising that listeners are less sensitive to intonational mismatch than to durational mismatch.

It may be argued that the delay observed for the constructed singular forms is not the result of intonational mismatch, but instead of the fact that the signal for the constructed singulars has been manipulated whereas the signal for the normal singulars has not been manipulated. We cannot rule out this possibility. However, the fact that the constructed singular forms sounded extremely natural suggests to us that intonational mismatch does indeed have a role to play, even though the 10 ms effect observed here may constitute an upper limit for the effect of intonational mismatch for materials presented in isolation. Subsequent research is needed to elucidate the potential effects of intonational information in the speech signal.

In all experiments so far we employed a number decision task. In the next and last experiment, we replicate the basic finding using another experimental paradigm, auditory lexical decision. We opted for lexical decision for two reasons. First, auditory lexical decision is a task in which the number of syllables is irrelevant: Whereas for number decision the number of syllables, and thus the prosodic structure of the stem, is informative with respect to the decision to be made, for lexical decision it is not. A first question addressed by Experiment 2.4 therefore is whether listeners are also sensitive to prosodic cues under these circumstances. Second, the responses to normal and constructed pseudoword singulars may shed light on whether the prosodic mismatch effect observed for existing words results purely from the representations stored in the mental lexicon or whether it is mediated at some pre-lexical level.

Experiment 2.4

Method

Participants. Forty-two participants, mostly students at the University of Nijmegen, were paid to participate in the experiment. All were native speakers of Dutch. None of them had participated in Experiments 2.1 to 2.3.

Materials. Four experimental item types were included in the experiment: normal and constructed singular *word* items, and normal and constructed singular *pseudoword* items. The word items were the exact experimental items as used in Experiment 2.1 (i.e., 48 normal singular forms and 48 constructed singular forms).

Out of the singular word items, 48 singular pseudoword items were created by changing one to three phonemes in such a way that the phonotactic constraints of Dutch were not violated, and that the pseudowords' prosodic structure was identical to that of the words. Subsequently, the 'plural' forms of these pseudowords were created by adding the plural suffix *-en* [ə(n)], which is the appropriate allomorph as the stems consisted of a single syllable. The 48 singular and 48 plural forms were assigned to separate reading lists. The orders within these lists were randomized twice, resulting in 4 reading lists. Due to an error, one pseudoword eventually had to be removed from the design. The remaining 47 pseudowords are listed in Appendix B.

Additionally, 100 filler words were included in the experiment: 25 monomorphemic, uninflected nouns, 25 inflected nouns (plural and diminutive inflections), 25 uninflected and inflected verbs, and 25 uninflected and inflected adjectives. The number of syllables of the filler words ranged from one to three. Out of these filler words, 100 filler pseudowords were created by changing one to three phonemes, again in such a way that the phonotactic constraints of Dutch were not violated, and that the pseudowords' prosodic structure was identical to that of the words. The filler words and the filler pseudowords were assigned to one reading list. The order within this list was randomized three times, resulting in three reading lists.

One more reading list was created consisting of 10 words, 5 'singular' pseudowords, and 5 'plural' pseudowords. These items were used to create practice trials. The order within this list was randomized twice, resulting in two reading lists.

All 9 reading lists were recorded by the same native female speaker of Dutch as in the previous experiments. The recordings were made in a soundproof recording booth and subsequently digitized at 16 kHz.

From the reading lists containing the experimental pseudoword items, the best realizations (of two) of the singular and the plural forms were selected. The singular forms served as the normal singular pseudoword items. Constructed singular pseudoword items were created by splicing the 'stems' out of the plural forms. From the reading lists containing the filler items, the best realizations (of three) of all filler words and of all filler pseudowords were selected. Finally, from the lists containing the practice items, the best realizations (of two) of all words and pseudowords were selected. The 10 words and the 5 singular pseudowords were spliced out of the lists exactly as they were realized by the speaker. From the plural pseudowords we only selected the stems, creating 5 constructed singular pseudowords.

Three experimental trial lists and their complements were created in such a way that a given list contained 100 filler words, 100 filler pseudowords, 24 normal singular word items, 24 constructed singular word items, 24 (or 23) normal singular pseudoword items, and 23 (or 24) constructed singular pseudoword items. One list never contained both the normal and the constructed singular form of a single noun (word or pseudoword): If one list contained the normal singular form of a noun, then the constructed singular form of that noun was contained in its complementary list. The order of presentation of items was pseudo-randomized: No more than three singular forms of the same type occurred successively. Orders were identical in complementary lists. Participants were randomly assigned to experimental trial lists. The twenty practice trials were presented prior to the experiment.

The pseudoword items showed differences in duration and in intonation, similar to those observed in the word items (see Experiment 2.1): The normal singular forms were significantly longer (94 ms on average) than the constructed singular forms ($t(46) = 21.5, p < 0.0001$), and the constructed singular forms had a higher average fundamental frequency (5 Hz on average, 198 Hz for the normal forms and 203 Hz for the constructed forms) than the normal singular forms ($t(46) = -1.9, p < 0.1$). The mean difference in vowel duration was 15 ms ($t(46) = 8.1, p < 0.0001$), the mean difference in closure duration was 16 ms ($t(46) = 6.0, p < 0.0001$), and the mean difference in release noise duration was 71 ms ($t(46) = 26.1, p < 0.0001$). Table 2.6 lists the mean durations for the normal and constructed singular pseudowords. The magnitudes of the differences between the normal and the constructed forms were similar for words and pseudowords (duration: $F(1, 93) = 1.5, p = 0.22$; fundamental frequency: $F(1, 93) = 0.2, p = 0.64$).

Table 2.6: Experiment 2.2 – Mean durations (in ms) with *SD* for normal and constructed singular pseudowords.

	Normal singular form		Constructed singular form		Duration difference
	Duration	SD Duration	Duration	SD Duration	
Whole form	451	66	357	67	94
Vowel	124	42	109	42	15
Closure	93	21	77	15	16
Release noise	108	23	37	12	71

Procedure. Participants were instructed to decide as quickly as possible whether the form they heard was a word or a pseudoword. They responded by pressing one of two buttons on a button box. Each trial consisted of the presentation of a warning tone (377 Hz) for 500 ms, followed after an interval of 450 ms by the auditory stimulus. Stimuli were presented through Sennheiser headphones. Reaction times were measured from stimulus offset. Each new trial was initiated 2500 ms after offset of the previous stimulus. When a participant did not respond within 2000 ms post-offset, a time-out response was recorded. Prior to the actual experiment, the set of practice trials was presented, followed by a short pause. Two short pauses were included in the experiment, resulting in three experimental trial blocks of approximately equal size. The total duration of the experimental session was approximately 30 minutes.

Results and discussion

The data of all participants were included in the analyses, since they all showed error rates below 20%. Nine word items and three pseudoword items elicited error rates above 20%. These items and their corresponding forms in the complementary condition were excluded from the analyses. Table 2.7 lists the mean reaction times (calculated over the correct trials only) and the percentages of incorrect trials for the four experimental item types (after exclusion of the items with high error rates).

Table 2.7: Experiment 2.4 – Mean response latencies (in ms) measured from word offset (calculated over correct trials only) with *SD* and percentages of incorrect trials for normal and constructed singular word items and for normal and constructed singular pseudoword items.

Item type	RT	SD	% Incorrect
Normal singular word item	442	79	2.3
Constructed singular word item	531	89	3.3
Normal singular pseudoword item	524	87	2.0
Constructed singular pseudoword item	583	79	4.3

The reaction times to the constructed experimental forms were significantly longer (89 ms on average for words, 59 ms on average for pseudowords) than the reaction times to the normal experimental forms ($F1(1,41) = 100.4, p < 0.0001; F2(1,81) = 55.6, p < 0.0001$; no interaction of type of singular form by word status: $F1(1,41) = 1.7, p = 0.22; F2(1,81) = 2.3, p = 0.14$). In order to rule out the possibility that the observed delay to the constructed singular forms is solely the result of the splicing manipulation applied to these forms, we ran a covariance model on the reaction time data for the constructed singular forms. A linear model was fitted to the data of each participant separately (cf. Lorch & Myers, 1990), in which log reaction times were predicted by the duration of the form itself, by the durational difference score (i.e., the difference in duration between the normal and the constructed form), and by lexical status (word versus pseudoword). T-tests on the coefficients of the subjects on the three predictor variables yielded a facilitatory main effect of duration ($t(41) = -8.6, p < 0.0001$) and a significant interaction of durational difference by lexical status ($t(41) = -4.9, p < 0.0001$). A multi-level extension of the Lorch and Myers technique (Pinheiro & Bates, 2000) revealed that for words, durational difference had an inhibiting effect: the larger the durational difference, the longer the reaction times ($t(1815) = -3.2, p < 0.01$). For pseudowords, however, we obtained the opposite effect: the larger the durational difference, the shorter the reaction times ($t(1815) = 2.11, p < 0.05$). In other words, large prosodic (durational) mismatch appears to make words less word-like and pseudowords more pseudoword-like. A comparison between words and pseudowords of the coefficients for the correlation between durational differences and reaction times revealed that this correlation was significantly stronger for words than for pseudowords ($Z = -2.3, p < 0.05$).

To conclude, the results of this experiment show that the prosodic mismatch effect is not restricted to the number decision task, but is also visible in auditory lexical decision. It is clear that the participants took the prosodic cues into account, even though these cues were irrelevant for making auditory lexical decisions. Interestingly, the correlational analysis revealed that the prosodic mismatch effect was stronger for words than for pseudowords, suggesting a word-specific component to the prosodic mismatch effect.

General discussion

In this study, we investigated whether uninflected and inflected forms have different prosodic characteristics, and whether such characteristics are functional for the listener in distinguishing these forms, by reducing the ambiguity between them. We found that indeed such acoustic differences exist between uninflected and inflected forms, and that listeners are sensitive to them. When prosodic information mismatches segmental information, participants show a delay in processing (Experiment 2.1, 2.2, and 2.3, number decision, and Experiment 2.4, auditory lexical decision). We refer to this phenomenon as the *prosodic mismatch effect*. In distinguishing singular forms from the stems of their corresponding plural forms, two sources of non-segmental information in particular play an important part: duration (Experiment 2.1, 2.2, and 2.4) and intonation (Experiment 2.3). The acoustic mismatch effect occurs both in singulars and in plurals (Experiment 2.2), and in words and pseudowords (Experiment 2.4). The prosodic differences between uninflected forms and the stems of their corresponding inflected forms reduce the ambiguity between these forms. Our results suggest that these acoustic cues help the perceptual system in determining early in the signal whether an inflected (bisyllabic) or an uninflected (monosyllabic) form is heard.

The existence of the prosodic mismatch effect has important consequences for theories of lexical processing and lexical representation. In classical models of lexical processing, the dominant view has been that all phonetic variation in the speech signal is abstracted away from through acoustic-phonetic analysis, in which the speech signal is translated into a string of discrete phoneme-like units. This abstract string constitutes an intervening representational level through which the speech signal is mapped onto representations in the mental lexicon (Pisoni & Luce, 1987). Since the abstract segmental representation of the singular form would be identical to that of the stem of the plural form, there is no reason why a delay in processing would occur when there is a mismatch between prosodic and segmental information: After acoustic-phonetic analysis, the processing system no longer has access to prosodic information, neither at the pre-lexical level, nor at the lexical level. Thus, models of speech perception that propose a strictly phonemic account of lexical access are challenged by the acoustic mismatch effect observed in the present study.

An alternative account of lexical processing and representation, originally proposed as an answer to the inability of the conventional models to deal with phonological variation, abandons the notion of an intervening segmental level (Lahiri & Marslen-Wilson, 1991). Instead, it assumes that the input to the lexical level is featural. It furthermore assumes that there is a single phonological underlying representation for each lexical item, which abstracts away from all surface detail, and which is compatible with all phonologically permissible variants in a given context. The lexical representations in this framework contain only distinctive and marked information. Predictable information is not specified. For instance, in English a word-final /n/ can be realized as /n/, as /m/, or as /ŋ/, depending on the place of articulation of the following segment: *green berry* (/m/), *green glass* (/ŋ/) versus *green dress* (/n/). Hence, the final nasal of *green* is unspecified for place of articulation. In other words, in this framework, phonemic variation is not represented lexically if it is predictable. This suggests that predictable variation that is prosodic in nature is not represented lexically either. If so, it is unclear how the prosodic mismatch effect might arise in this kind of approach.

An approach which can account for the prosodic mismatch effect is that of Johnson (1997). He trained a connectionist (exemplar-based) model on vector quantized speech data, which contained — among other things — information regarding the durations of the segments. Johnson's model correctly anticipated whether the incoming syllable was followed by another (unstressed) syllable or not. Davis et al. (2002) also favor a subsymbolic model that is sensitive to subphonemic properties of the acoustic input.

Our explanation for the occurrence of the prosodic mismatch effect is framed in the exemplar-based or episodic approach of Goldinger (1998), but it can be incorporated in other theoretical approaches as well. We think that in parallel to the processing of the acoustic signal of the stem, an expectation regarding the number of unstressed syllables that will follow is built up based on the durations of the segments. A delay in processing will occur when this expectation is violated by the segmental material that either does or does not follow the stem. The build-up of an expectation regarding the possible continuation of the signal would be advantageous at several levels.

First, it would provide information regarding the prosodic make-up of the utterance. Salverda et al. (2003) point out that subtle acoustic cues may signal the presence or absence of a prosodic word boundary. They argue that a prosodic representation is computed, based in part on these acoustic cues and in parallel to the segmental encoding. This prosodic representation would contribute to lexical activation by favoring candidates whose boundaries are aligned with the hypothesized prosodic boundary.

Furthermore, the expectation about whether an unstressed syllable is to follow would also provide information regarding the morphological make-up of the incoming speech signal. The prosodic cues signal whether the acoustic signal at hand is that of an unmodified (monosyllabic) stem or that of the same stem but now followed by an unstressed (inflectional or derivational) suffix or by an (unstressed) clitic. We showed that listeners probably determine whether a stem is part of a morphologically simplex form or not, well before the segmental information comes in that signals the presence or absence of a suffix (or clitic).

If it is true that the prosodic mismatch effect arises from the violation of an expectation that is based on the durations of segments, then the question arises how it is possible that listeners are sensitive to these durations, given the enormous variability in the temporal structure of speech. Speech rate varies between speakers, within speakers, and within speakers even within one sentence. Hence, the absolute durations of segments will vary tremendously from utterance token to utterance token. We think that the solution of this riddle lies in the relative durations of the segments in the stem.

Consider Figure 2.5, which summarizes the distributions of durations by means of boxplots of the onset, the vowel, and the coda of the monosyllabic stems of the words from Experiment 2.1 (upper panel) and pseudowords from Experiment 2.4 (lower panel). The boxes show the interquartile range, the horizontal line in the box denotes the median, and the 'whiskers' extend to the observations within 1.5 times the interquartile range. Outliers beyond this range are represented by individual circles. Differences in duration that are significant in two-tailed pairwise *t*-tests as well as in two-tailed paired Wilcoxon tests ($p < 0.0001$) are marked with asterisks.

What Figure 2.5 shows is that there is no reliable difference in duration between the onset of the singular form and the onset of the stem of the corresponding plural form. For the pairs of onsets of existing words, there is a 7 ms difference that fails to reach significance ($t(47) = 1.7, p = 0.10$). For the onset pairs in pseudowords, there is a 8 ms difference in the opposite direction (the onsets of stems in plurals tend to be longer than those of singulars) that also does not reach significance

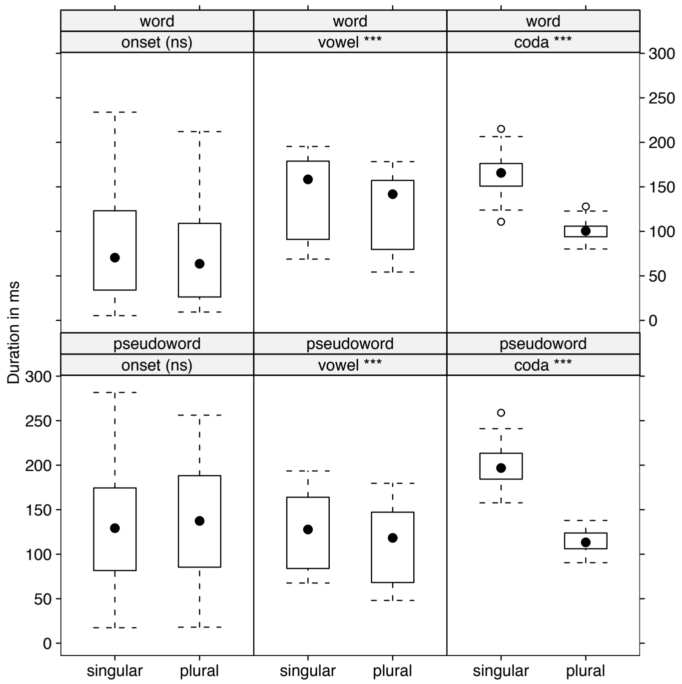

Figure 2.5: Duration in ms of the onset, the vowel, and the coda in singulars (sg) and in the stems of plurals (pl), for words (upper panel) and pseudowords (lower panel).

($t(46) = -1.5, p = 0.14$).

These small and non-significant differences in duration of the onset contrast with the longer and very significant difference in duration for the vowels (17 ms for the words and 15 ms for the pseudowords). For the codas, the difference in duration is even greater (63 ms for the words and 87 ms for pseudowords, most of which is due to the release noise duration of the final plosive). Considered jointly, this pattern of results suggests that the duration of the onset is a stable anchor point against which the duration of the vowel as well as the duration of the coda can be calibrated. If the durations of vowel and coda compared to that of the onset are relatively long, the incoming speech signal is likely to be a singular. If these durations are relatively short, the likelihood increases that it will be part of a morphological continuation form. In other words, we think that the relative durations of vowel and coda with respect to the onset provide the acoustic information that in our experiments gives rise to the prosodic mismatch effect.

Relative durations differ from word to word. For instance, the relative duration of the vowel with respect to the onset will depend on whether the vowel is phonemically long or short, as well as on the number of segments in the onset. Similarly, the relative length of the coda varies with the number of segments in the coda and in the onset. In addition, specific combinations of segments in the syllable may affect their duration (Waals, 1999). We therefore hypothesize that the relevant information is provided lexically, with a given lexical form, in our experiment a given singular or its plural, having a prototypical distribution of relative segmental durations. In other words, we propose that a lexical entry does not only specify the segments and their order, but also the relative durations of vowel and coda with respect to the onset. (In the subsymbolic approach of Johnson (1997), the item-specific distributions would presumably be coded in the weights of the connections in the network mapping vector-quantized speech input onto lexical representations.) This view is consistent with the finding that the correlation between prosodic (durational) mismatch and reaction times was stronger for words than for pseudowords, suggesting item-specific support for the prosody-based expectation regarding the number of syllables to follow for existing words.

The prosodic mismatch effect for pseudowords (Experiment 2.4) points to the existence of a general rule or of an analogical mechanism for building up an expectation of whether an unstressed syllable will follow, as no lexical entries are available for pseudowords. Given an analogical mechanism that generalizes over stored exemplars, the prosodic mismatch effect in pseudowords can be viewed as resulting from implicit knowledge of prosodic structure that emerges from the patterns that are present in the lexicon. In a subsymbolic framework, the prosodic mismatch effect for pseudowords would reflect the implicit generalizations of the network with respect to the co-occurrences of segmental durations and syllable structure. In more general terms, the prosodic mismatch effect for pseudowords probably reflects the unconditional probabilities for the co-occurrences of segmental durations and syllable structure. In the case of words, these unconditional probabilities might be supplemented by conditional probabilities based on the co-occurrences of the sequence of segments constituting a word's form representation, the durations of these segments, and their syllable structure. The hypothesis that durational structure is part of the lexical representations of words is compatible with Goldinger's (1998) episodic (or exemplar-based) theory, according to which experience with spoken word tokens leaves detailed traces of these tokens in memory. It is also compatible with the linguistic distributional evidence brought together by Bybee (2001), evidence which shows that phonologically redundant information is stored in the (mental) lexicon. Furthermore, it is compatible with Pierrehumbert's exemplar-based framework (2002), in which each individual word has an associated probability distribution (exemplar cloud) for each of its segments.

The importance of durational information is also supported by the pattern of frequency effects in our experiments, a pattern which strongly suggests that the durational information in the stem codetermines which of two representations (singular or plural) becomes most active. For all experiments, we conducted multi-level covariance analyses (Pinheiro & Bates, 2000) in which reaction times were predicted by Duration, Singular Surface Frequency, Plural Surface Frequency, and — where applicable — Durational Difference (between normal and constructed form). We will only discuss the effects of Singular Surface Frequency and of Plural Surface Frequency here.

In the number decision experiments, we observed effects of Singular Surface Frequency in all cases *except* when both the segmental and the durational information pointed to a plural form (i.e., in the case of the normal plural forms in Experiment 2.2). In other words, if either source of information (segmental or durational) in the acoustic signal points to a singular form, the singular representation is activated, even when there is a mismatch between the different sources of information in the signal [3]. Plural Surface Frequency, on the other hand, has an effect whenever the durational information points to the plural form, irrespective of what form the segmental information points to (i.e., in the case of the constructed singular forms in Experiment 2.1 and the normal plural forms in Experiment 2.2) [4]. In other words, in a number decision task, the durational information in the stem appears to codetermine whether the singular or the plural representation is activated:

[3] In Experiment 2.1 (prosodic — durational and intonational — difference between normal and constructed singular forms), Singular Surface Frequency had a facilitatory effect on reaction times to both the normal singular forms (with segmental and prosodic cues pointing to the singular; $t(1044) = -2.7, p < 0.01$) and the constructed singular forms (with segmental cues pointing to the singular but prosodic cues pointing to the plural; $t(1034) = -3.8, p < 0.001$). The higher the Singular Surface Frequency, the easier it was for participants to give the response 'singular' to both the normal and the constructed singular forms. In Experiment 2.2 (prosodic — durational and intonational — difference between normal and constructed plural forms), we observed a facilitatory effect of Singular Surface Frequency ($t(998) = -2.9, p < 0.01$) for the constructed plural forms only (i.e., for the forms that carried the prosodic characteristics of the singular). In Experiment 2.3, in which the two types of singular forms differed in intonation, but not in duration (and in fact both carry the durational characteristics of the singular), we observed a facilitatory effect of Singular Surface Frequency for both normal and constructed singular forms ($t(2269) = -2.8, p < 0.01$).

[4] In Experiment 2.1, Plural Surface Frequency had an inhibiting effect on the reaction times to the constructed singular forms only (i.e., to the forms that carried the prosodic — durational and intonational — characteristics of the plural). The higher the Plural Surface Frequency, the more *difficult* it was for participants to give the response 'singular' to the constructed singular forms. In Experiment 2.2, we observed a facilitatory effect of Plural Surface Frequency for the normal plural forms ($t(999) = -2.7, p < 0.01$). In Experiment 2.3 (intonational but no durational difference), there was no effect of Plural Surface Frequency, neither for the normal forms nor for the constructed forms ($t(2269) = 0.7, p = 0.50$). This latter finding shows that only the presence of intonational cues to a particular form is not sufficient to activate that form.

Durational cues to the plural form lead to activation of the plural representation, durational cues to the singular form lead to activation of the singular representation.

When in a number decision experiment, segmental information points to a singular form whereas durational information points to a plural form (i.e., in the case of the constructed singular forms in Experiment 2.1), we observe competition between the singular and the plural form: Both the singular and the plural representations are activated. In the normal case (i.e., in the case of the normal singular forms in Experiment 2.1 and in the case of the normal plural forms in Experiment 2.2), no competition is observed: Only the correct representations are activated. The ambiguity between the singular and the plural form appears to be resolved through the durational differences in the stem. This finding reduces the competition problem that is the result of having stored lexical representations for inflected forms in lexical memory. Given the prosodic differences documented in this study, the inflected form might well be a less strong cohort competitor for the uninflected form and vice versa.

In the lexical decision experiment (Experiment 2.4), we observed a different pattern of frequency effects. There were facilitatory effects of both Singular Surface Frequency and Plural Surface Frequency, for both normal and constructed singular forms [5]. We observed no competition, contrary to in the number decision experiments. Interestingly, for lexical decision, the relevant information is whether the perceived segments form an existing word. As the distinction between the singular and the plural is irrelevant in lexical decision, the support for the singular and plural is pooled: Both the singular and the plural representations support a positive lexical decision.

The prosodic mismatch effect documented in this study has important consequences for our understanding of the morphological structure of complex words. The way words are written in languages such as Dutch and English suggests that they consist of stems and affixes that are strung together as beads on a string. Phonemic transcriptions convey the same impression. Our experiments show that this impression is wrong. Plurals are not just singulars with an additional suffix. The precise acoustic realization of the stem provides crucial information to the listener about the morphological context in which the stem appears.

[5] In Experiment 2.4 (lexical decision), we observed facilitatory effects of both Singular Surface Frequency and Plural Surface Frequency for normal singular forms (Singular Surface Frequency: $t(874) = -3.7, p < 0.001$, one-tailed; Plural Surface Frequency: $t(874) = -2.7, p < 0.01$, one-tailed) and for constructed singular forms (Singular Surface Frequency: $t(864) = -1.6, p = 0.05$, one-tailed; Plural Surface Frequency: $t(864) = -1.8, p < 0.05$, one-tailed). (We applied one-tailed tests as frequency effects are always facilitatory for lexical decision.)

References

Alegre, M. and Gordon, P.: 1999, Frequency effects and the representational status of regular inflections, *Journal of Memory and Language* **40**, 41–61.

Aronoff, M.: 1976, *Word Formation in Generative Grammar*, MIT Press, Cambridge, Mass.

Baayen, R. H., Burani, C., and Schreuder, R.: 1997, Effects of semantic markedness in the processing of regular nominal singulars and plurals in Italian, *in* G. E. Booij and J. v. Marle (eds), *Yearbook of Morphology 1996*, Kluwer Academic Publishers, Dordrecht, pp. 13–34.

Baayen, R. H., Dijkstra, T., and Schreuder, R.: 1997, Singulars and plurals in Dutch: Evidence for a parallel dual route model, *Journal of Memory and Language* **37**, 94–117.

Baayen, R. H., McQueen, J., Dijkstra, T., and Schreuder, R.: 2003, Frequency effects in regular inflectional morphology: Revisiting Dutch plurals, *in* R. H. Baayen and R. Schreuder (eds), *Morphological structure in language processing*, Mouton de Gruyter, Berlin, pp. 355–390.

Baayen, R. H., Piepenbrock, R., and Van Rijn, H.: 1993, *The CELEX lexical database (CD-ROM)*, Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Baayen, R. H., Schreuder, R., De Jong, N. H., and Krott, A.: 2002, Dutch inflection: the rules that prove the exception, *in* S. Nooteboom, F. Weerman, and F. Wijnen (eds), *Storage and Computation in the Language Faculty*, Kluwer Academic Publishers, Dordrecht, pp. 61–92.

Bertram, R., Laine, M., Baayen, R. H., Schreuder, R., and Hyönä, J.: 1999, Affixal homonymy triggers full-form storage even with inflected words, even in a morphologically rich language, *Cognition* **74**, B13–B25.

Boersma, P. and Weenink, D. J. M.: 1996, PRAAT, a system for doing phonetics by computer, version 3.4, *Institute of Phonetic Sciences of the University of Amsterdam, Report* **132**, Available at http://www.fon.hum.uva.nl/praat/.

Booij, G. E.: 1995, *The phonology of Dutch*, Clarendon Press, Oxford.

Bybee, J. L.: 2001, *Phonology and language use*, Cambridge University Press, Cambridge.

Cambier-Langeveld, T.: 2000, *Temporal marking of accents and boundaries*, LOT, Amsterdam.

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G.: 2002, Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition, *Journal of Experimental Psychology: Human Perception and Performance* **28**, 218–244.

Dressler, W. U., Mayerthaler, W., Panagl, O., and Wurzel, W. U.: 1987, *Leitmotifs in Natural Morphology*, Benjamins, Amsterdam.

Fowler, C. A.: 1977, *Timing control in speech production*, Indiana University Linguistics Club, Bloomington, Indiana.

Goldinger, S. D.: 1998, Echoes of echoes? An episodic theory of lexical access, *Psychological Review* **105**, 251–279.

Johnson, K.: 1997, The auditory/perceptual basis for speech segmentation, *Ohio State University Working Papers in Linguistics* **50**, 101–113.

Lahiri, A. and Marslen-Wilson, W. D.: 1991, The mental representation of lexical form: A phonological approach to the recognition lexicon, *Cognition* **38**, 245–294.

Lehiste, I.: 1972, *Suprasegmentals*, MIT Press, Cambridge, Mass.

Lindblom, B. and Rapp, K.: 1973, Some temporal regularities of spoken Swedish, *PILUS* **21**, 1–59.

Lorch, R. F. and Myers, J. L.: 1990, Regression analyses of repeated measures data in cognitive research, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **16**, 149–157.

Marslen-Wilson, W. D.: 1990, Activation, competition, and frequency in lexical access, *in* G. T. M. Altmann (ed), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, The MIT Press, Cambridge, Mass, pp. 148–172.

Marslen-Wilson, W. D., Moss, H. E., and Van Halen, S.: 1996, Perceptual distance and competition in lexical access, *Journal of Experimental Psychology: Human Perception and Performance* **22**, 1376–1392.

McCarthy, J. J. and Prince, A.: 1993, Generalized alignment, *in* G. E. Booij and J. v. Marle (eds), *Yearbook of Morphology*, Kluwer Academic Publishers, Dordrecht, pp. 79–154.

McClelland, J. L. and Elman, J. L.: 1986, The TRACE model of speech perception, *Cognitive Psychology* **18**, 1–86.

Munhall, K., Fowler, C., Hawkins, S., and Saltzman, E.: 1992, ”Compensatory shortening” in monosyllables of spoken English, *Journal of Phonetics* **29**, 225–239.

Nooteboom, S. G.: 1972, *Production and perception of vowel duration: A study of the durational properties of vowels in Dutch*, Unpublished doctoral dissertation, University of Utrecht, Utrecht.

Norris, D. G.: 1994, Shortlist: A connectionist model of continuous speech recognition, *Cognition* **52**, 189–234.

Pierrehumbert, J.: 2002, Word-specific phonetics, *in* C. Gussenhoven and N. Warner (eds), *Laboratory Phonology VII*, Mouton de Gruyter, Berlin, pp. 101–140.

Pinheiro, J. C. and Bates, D. M.: 2000, *Mixed-effects models in S and S-PLUS*, Statistics and Computing, Springer, New York.

Pisoni, D. B. and Luce, P. A.: 1987, Acoustic-phonetic representation in word recognition, *Cognition* **25**, 21–52.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., and Gremmen, F.: 1999, How to deal with ‘the language-as-fixed-effect-fallacy’: Common misconceptions and alternative solutions, *Journal of Memory and Language* **41**, 416–426.

Salverda, A., Dahan, D., and McQueen, J.: 2003, The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension, *Cognition* **90**, 51–89.

Schreuder, R., De Jong, N. H., Krott, A., and Baayen, R. H.: 1999, Rules and rote: beyond the linguistic either-or fallacy, *Behavioral and Brain Sciences* **22**, 1038–1039.

Sereno, J. and Jongman, A.: 1997, Processing of English inflectional morphology, *Memory and Cognition* **25**, 425–437.

Waals, J.: 1999, *An experimental view of the Dutch syllable*, Holland Academic Graphics, The Hague.

Appendix A

The experimental nouns used in Experiment 2.1:

Singular	Plural	Singular	Plural
1. beek	beken	25. lat	latten
2. boot	boten	26. lip	lippen
3. bout	bouten	27. map	mappen
4. breuk	breuken	28. noot	noten
5. brok	brokken	29. pet	petten
6. buik	buiken	30. peuk	peuken
7. dijk	dijken	31. plaat	platen
8. draak	draken	32. pruik	pruiken
9. duit	duiten	33. rat	ratten
10. feit	feiten	34. reep	repen
11. geit	geiten	35. rok	rokken
12. graat	graten	36. schaap	schapen
13. grap	grappen	37. spreuk	spreuken
14. grot	grotten	38. straat	straten
15. heup	heupen	39. struik	struiken
16. kaak	kaken	40. taak	taken
17. kip	kippen	41. tak	takken
18. klip	klippen	42. vak	vakken
19. knaap	knapen	43. vlok	vlokken
20. knop	knoppen	44. wet	wetten
21. krat	kratten	45. wrak	wrakken
22. kruik	kruiken	46. wrat	wratten
23. kuit	kuiten	47. zaak	zaken
24. lap	lappen	48. zweep	zwepen

Appendix B

The pseudowords used in Experiment 2.4:

Singular	Plural	Singular	Plural
1. beep	bepen	25. kaat	katen
2. bijk	bijken	26. knaat	knaten
3. brek	brekken	27. paak	paken
4. breut	breuten	28. peut	peuten
5. draap	drapen	29. plaak	plaken
6. fap	fappen	30. plik	plikken
7. feik	feiken	31. rak	rakken
8. fek	fekken	32. schoet	schoeten
9. fip	fippen	33. soot	soten
10. fnok	fnokken	34. sprek	sprekken
11. foot	foten	35. strat	stratten
12. frap	frappen	36. stroek	stroeken
13. fruik	fruiken	37. suik	suiken
14. gaak	gaken	38. tek	tekken
15. get	getten	39. trak	trakken
16. geup	geupen	40. trit	tritten
17. glit	glitten	41. veek	veken
18. gop	goppen	42. weip	weipen
19. gouk	gouken	43. wop	woppen
20. graak	graken	44. wot	wotten
21. grat	gratten	45. wuik	wuiken
22. grok	grokken	46. zaap	zapen
23. gruik	gruiken	47. zwoep	zwoepen
24. guik	guiken		

Prosodic cues for morphological complexity in Dutch and English

CHAPTER 3

This chapter will appear as Rachèl J. J. K. Kemps, Lee H. Wurm, Mirjam Ernestus, Robert Schreuder, and R. Harald Baayen: Prosodic cues for morphological complexity in Dutch and English, *Language and Cognitive Processes*.

Abstract

Previous work has shown that Dutch listeners use prosodic information in the speech signal to optimize morphological processing: Listeners are sensitive to prosodic differences between a noun stem realized in isolation and a noun stem realized as part of a plural form (in which the stem is followed by an unstressed syllable). The present study, employing a lexical decision task, provides an additional demonstration of listeners' sensitivity to prosodic cues in the stem. This sensitivity is shown for two languages that differ in morphological productivity: Dutch and English. The degree of morphological productivity does not correlate with listeners' sensitivity to prosodic cues in the stem, but it is reflected in differential sensitivities to the word-specific log odds ratio of encountering an unshortened stem (i.e., a stem in isolation) versus encountering a shortened stem (i.e., a stem followed by a suffix consisting of one or more unstressed syllables). In addition to being sensitive to the prosodic cues themselves, listeners are also sensitive to the probabilities of occurrence of these prosodic cues.

Introduction

In languages with a concatenative morphological system, such as Dutch and English, morphologically complex words consist of (combinations of) stems preceded by one or more prefixes and/or followed by one or more suffixes. The orthographic representations of morphologically complex words suggest that these stems, prefixes, and suffixes are strung together as beads on a string. Acoustically, however, the realizations of morphemes that are concatenated to form a morphologically complex word are different from the realizations of these morphemes when produced in isolation, even when the morphemes are phonemically unchanged after concatenation. One of the reasons for this is that, in stress-timed languages, the duration of a stressed vowel reduces as a function of the number of unstressed syllables that follow (Nooteboom, 1972, for Dutch; Lehiste, 1972, Fowler, 1977, for English; Lindblom & Rapp, 1973, for Swedish). In other words, the duration of the vowel in a syllable is shorter when this syllable is followed by one or more unstressed syllables than when it is produced in isolation. For example, the vowel in the first syllable of *walking* is shorter than the vowel in *walk*.

Previous studies have shown that listeners are very sensitive to such acoustic differences. It has been shown that listeners can use these differences as cues to distinguish strings that are initially phonemically ambiguous between a word and a morphologically unrelated continuation form of that word. Salverda, Dahan, and McQueen (2003) recorded participants' eye movements while they listened to Dutch sentences including a word with a morphologically unrelated onset-embedded word (e.g., *hamster* containing *ham*). The participants saw four pictures of objects on a computer screen and were instructed to use the computer mouse to move the picture of the object that was mentioned in the sentence. There were more fixations to a picture representing the embedded word (*ham*) when the first syllable of the target word (*hamster*) had been replaced by a recording of the embedded word than when it came from a different recording of the target word. This demonstrates that segmentally ambiguous sequences can contain acoustic cues (in this case, the duration of the embedded word *ham* relative to the duration of its corresponding syllable in the target word *hamster*), that modulate its lexical interpretation.

Similar results were obtained by Davis, Marslen-Wilson, and Gaskell (2002). In a gating task, participants were presented with sentence fragments. In one condition (long-word condition), the sentence fragments ended in a long carrier word of which the initial syllable formed an onset-embedded word (e.g., *captain* containing *cap*).

In the other condition (short-word condition), the fragments ended in the short word corresponding to the initial syllable of the carrier word followed by a word with an onset that matched the continuation of the longer carrier word (e.g., *cap* **t***ucked* versus *cap***t***ain*). The first syllable in the short-word condition was significantly longer than the first syllable in the long-word condition, and there was a marginally significant difference in average fundamental frequency (average fundamental frequency was higher in the long-word condition than in the short-word condition). Significantly more short-word responses were made to gates from short-word stimuli than to gates from long-word stimuli, suggesting that listeners take advantage of the acoustic differences that exist between short and long word sequences. Similar results were obtained in a cross-modal priming task. The stimuli from the gating task were presented up to the offset of the first syllable of the target word (e.g., *cap* from either *cap* or *captain*) as auditory primes, and were followed by a visual target that was either the short word (*cap*) or the long word (*captain*). Greater facilitation occurred when prime syllables came from the same word as the target.

More recently, it has been shown that listeners are also sensitive to acoustic differences between phoneme strings that are initially ambiguous between a stem and a morphologically *related* continuation form of that stem, in particular, between a singular and a plural form of a noun (Kemps, Ernestus, Schreuder, & Baayen, submitted; Chapter 2 of this thesis). In Dutch, the regular plural form of many nouns consists of the noun stem and the plural suffix *-en*, which is usually realized as just a schwa (e.g., *boek* [buk] 'book' – *boeken* [bukə] 'books'). As a result of the addition of the schwa, the stem of the plural form is durationally and intonationally different from the stem realized in isolation (the singular form). In what follows, we will refer to such non-segmental differences in duration and intonation as prosodic differences. Such differences partly reflect differences in syllable structure. For instance, in the plural *boe-ken* [bu-kə], the suffix *-en* [ə] induces resyllabification of the stem-final obstruent [k] as onset of the next syllable and, as a consequence, the stem vowel is syllable-final in the plural [bu-kə] as opposed to syllable-medial in the singular *boek* [buk]. Listeners were presented with singular forms and with stems that were spliced out of plural forms. These stimuli were segmentally identical, but the stems of the plural forms carried mismatching prosodic information: The absence of a plural suffix pointed to the singular form, whereas the prosodic information pointed to the plural form. When presented with the mismatching forms, listeners were significantly delayed, both in a number decision task as well as in a lexical decision task. Similar results were obtained when listeners were presented with plural forms of which the stems carried either matching or mismatching prosodic information (i.e., plurals of which the stems originated either from another token of the plural form or from a realization of the singular form), and also when listeners were presented with pseudowords of which the 'stems' carried either matching or mismatching prosodic information (i.e., pseudowords of which the stems were originally realized in isolation or in combination with a plural suffix). Importantly, the magnitude of this prosodic mismatch effect, that is, the magnitude of the delay in response latencies, correlated with the magnitude of the durational mismatch: The larger the durational difference between the stem realized in isolation and the stem realized as part of the plural form, the larger the delay. This correlation was stronger for words than for pseudowords.

The prosodic differences between uninflected forms and the stems of their corresponding inflected forms reduce the ambiguity between these forms. The observed sensitivity of listeners to these prosodic differences suggests that these acoustic cues help the perceptual system in determining early in the signal whether an inflected (bisyllabic) or an uninflected (monosyllabic) form is likely to be heard. Plurals are not singulars with an additional suffix. The precise acoustic realization of the stem provides crucial information to the listener about the morphological context in which the stem appears.

The present study, employing a lexical decision task, aims at replicating these findings for different types of morphologically complex forms in Dutch, and at extending the investigation of listeners' sensitivity to prosodic cues for morphological complexity to another language, English. The morphologically complex forms under investigation in the present study are comparatives (inflection) and agent nouns (derivation). Studying the effects of prosodic mismatch in the processing of stems of agent nouns and of comparatives in both Dutch and English enables us to determine whether the effects observed in the processing of singular and plural forms in Dutch are specific to plural formation in Dutch, or whether they generalize to a different type of inflection, to derivation, and to a different language.

In Dutch and English, many agent nouns are formed by adding the suffix *-er* (Dutch: [əʀ]; English: [ɚ]) to the stem, which is a verb stem. For example, the English agent noun *worker* [wɝkɚ] consists of the verb stem *work* [wɝk] and the deverbal agentive suffix *-er* [ɚ]. Similarly, the Dutch agent noun *werker* [wɛʀkəʀ] consists of the verb stem *werk* [wɛʀk] and the deverbal agentive suffix *-er* [əʀ]. The suffix *-er* is homonymous (see Booij, 1979, for the many meanings of the suffix *-er* in Dutch): Many comparatives are also formed by adding the suffix *-er* to the stem, which in this case is an adjective. Thus, the English comparative *fatter* [fætɚ] consists of the adjective *fat* [fæt] and the comparative suffix *-er* [ɚ]. The Dutch comparative *vetter* [vɛtəʀ] consists of the adjective *vet* [vɛt] and the comparative suffix *-er* [əʀ]. The affixation of the suffix *-er* leads to shortening of the preceding stem and to changes in syllable structure. In the present study, employing a lexical decision task, we investigated whether listeners are sensitive to such prosodic differences between monosyllabic stems and the stems of bisyllabic complex forms. We presented listeners with stems of agent nouns and comparatives that carried either matching or mismatching prosodic information. If listeners are sensitive to the prosodic cues in the stem, they are expected to be slowed down in their responses when there is a mismatch between the number of syllables on the one hand, and the prosodic information in the acoustic signal on the other hand. If not, in other words, if listeners attend to segmental information only, mismatching prosodic information should not affect response latencies. Note that information about the identity of the complex forms that the stems originated from was not available to our listeners. The stem *werk* ('work'), for instance, originating from the agent noun *werker* ('worker') could just as well have originated from the infinitive verbal form *werken* ('to work'). We were therefore not interested in potential effects of the type of complex form that the stems originated from, but purely in the question of whether the prosodic mismatch effect observed in earlier work would generalize to different materials, and to a different language.

Dutch and English differ in morphological richness, in particular in the number of continuation forms that are possible given a certain monomorphemic stem. For example, whereas the verbal inflectional paradigm of the Dutch word *wandelen* ('to walk') consists of the forms *wandel*, *wandelt*, *wandelen*, *wandelde*, *wandelden*, *gewandeld*, *wandelend*, and *wandelende*, the verbal inflectional paradigm of the English word 'walk' contains only *walk*, *walks*, *walked*, and *walking*. In other words, the stem *wandel* is followed by an unstressed syllable in five inflectional forms, whereas the stem *walk* is followed by an unstressed syllable in only one form. In general, the number of continuation forms in which a stem is followed by an unstressed syllable is considerably smaller in English than in Dutch: Besides the richer verbal paradigm, Dutch also exhibits prenominal contextual inflection of adjectives (which consists of the addition of a schwa to the stem, e.g., *een groot boek* 'a big book' (neuter gender) versus *een grote auto* 'a big car' (common gender)), whereas English does not. Furthermore, in Dutch, most noun inflections consist of the addition of an unstressed syllable to the stem: Many plurals are formed by adding the suffix *-en* [-ən] to the stem. In English, on the other hand, many plurals are formed by adding the plural suffix *-s* ([s] or [z]) to the stem (no additional syllable, except for stems ending in sibilants). Finally, Dutch has more unstressed derivational suffixes than English. For example, diminutives in Dutch are formed by adding (an allomorph of) the diminutive suffix *-tje* [cə] to the stem, whereas diminutive derivation is not productive in English. It is conceivable that, as a consequence of these differences in the number of possible continuation forms in which a stem is followed by one or more unstressed syllables, Dutch and English listeners are not equally sensitive to prosodic cues in the stem that signal whether or not the stem will be followed by unstressed syllables. Possibly, English listeners are less sensitive to such prosodic cues, as, in English, a stem is relatively infrequently followed by an unstressed syllable.

We not only investigated the effect of prosodic mismatch on reaction times, but we also investigated the predictive value of two covariates that are word-specific indications of the prevalence of possible continuation forms: Syllable Ratio and Cohort Entropy.

Syllable Ratio gives a word-specific indication of the likelihood of observing an unshortened versus a shortened stem. It is defined as the log of the ratio which has as the numerator the Surface Frequency of a stem in isolation, and as the denominator the summed Surface Frequencies of words in which this stem is followed by an inflectional or derivational suffix consisting of one or more unstressed syllables (i.e., words in which the stem occurs in shortened form). We only considered inflectional and derivational suffixes that consist of one or more syllables containing schwa, so that the phonological shortening process in the stem is maximally comparable to that in the comparative stems and in the agent noun stems. For example, for the stem *strict*, the numerator of the Syllable Ratio would consist of the surface frequency of *strict* (i.e., 362), and the denominator would consist of the summed surface frequencies of *stricter*, *strictest*, and *strictness* (i.e., 69). All instances of the stem, irrespective of grammatical category, are included in the numerator of Syllable Ratio. Note that when the numerator is smaller than the denominator, the Syllable Ratio will be negative, as the log of reals between 0 and 1 is negative. Compounds were not included in the denominator, as little is known about phonological shortening within left constituents of compounds.

Syllable Ratio is the log odds ratio of observing an unshortened form versus observing a shortened form. All words occurred in monosyllabic form in the experiment. We therefore expected a facilitatory effect of Syllable Ratio: If Syllable Ratio was high for a given word (i.e., if a word occurs relatively often as a monosyllabic stem), faster response latencies were expected. A facilitatory effect of Syllable Ratio would constitute evidence for listener's sensitivity to the likelihood of occurrence of a certain prosodic manifestation of a particular stem.

Syllable Ratio only considers specific types of continuation forms, namely, the continuation forms that are morphologically related to the stem and in which the stem has undergone a shortening process as a result of the addition of one or more unstressed syllables. However, given a certain stem, many types of continuation forms are possible, including continuation forms that are not morphologically related. In order to rule out the possibility that an effect of Syllable Ratio is in fact just an effect of whatever is still present in the cohort at the final position in the stem, we need an index of the latter. We therefore introduce another covariate: the Cohort Entropy. Entropy is an information-theoretical measure, indicating the amount of uncertainty about the outcome of a selection process (Shannon, 1948, see also Moscoso del Prado Martín, Kostić, & Baayen, in press). Cohort Entropy (H) is defined as:

$$H = -\sum_{i=1}^{n} p_i \log p_i$$

in which p_i is the probability of a word given the n words that are still present in the cohort at the point in time when the stem-final segment of the target word has been perceived. In other words:

$$p_i = \frac{\text{Surface Frequency of Word}_i}{\text{Summed Surface Frequencies of } n \text{ Cohort Members at stem-final segment of target word}}$$

To illustrate, suppose that by the time that the final segment of Stem X has been perceived, the cohort consists of two word candidates: Word X_a and Word X_b. Word X_a has a surface frequency of 80 and Word X_b has a surface frequency of 20. For Stem Y, the stem-final cohort also consists of two word candidates (Word Y_a and Word Y_b), both of which have a surface frequency of 50. The Cohort Entropies for Stem X and Stem Y are calculated as follows (note that the Cohort Entropy is larger for Stem Y):

$$p_{X_a} = \frac{80}{80+20} = 0.80 \qquad p_{X_b} = \frac{20}{80+20} = 0.20$$

$$H_X = -\Big(0.80 * log(0.80) + 0.20 * log(0.20)\Big) = 0.50$$

$$p_{Y_a} = \frac{50}{50+50} = 0.50 \qquad p_{Y_b} = \frac{50}{50+50} = 0.50$$

$$H_Y = -\Big(0.50 * log(0.50) + 0.50 * log(0.50)\Big) = 0.69$$

Cohort Entropy is calculated at the stem-final segment as only stems (with either matching or mismatching prosodic information) were presented to our listeners. Included in the cohort are *all* possible continuation forms, that is, both morphologically related and morphologically unrelated continuation forms. For example, the cohort for the stem *bake* consists of *bake*, *bakes*, *baked*, *baking*, *baker*, *bakers*, *bakery*, *bakeries*, but also *bacon* and *bakelite*. Cohort Entropy is a non-phonologically and non-morphologically based measure, defined purely in terms of lexical competition. Note however that for monomorphemic stems (the type of stems used in the present study), morphologically related continuation forms (i.e., inflections, derivations, and compounds) are more prevalent than morphologically unrelated continuation forms, both type-wise and token-wise. (Counts are presented below.) We expect an inhibitory effect of Cohort Entropy: The more uncertainty, the longer the response latencies.

Experiment 3.1

Part A: Dutch

Method

Participants. Twenty participants, mostly students at the University of Nijmegen, were paid to participate in the experiment. All were native speakers of Dutch.

Materials. From the CELEX lexical database (Baayen, Piepenbrock, & Van Rijn, 1993) we selected all Dutch comparatives and agent nouns that contained a monomorphemic and monosyllabic stem, in which the stem ended in a voiceless plosive. In Dutch, underlyingly voiced obstruents are devoiced in syllable-final position and all stems realized in isolation therefore end in voiceless obstruents (final devoicing). The suffix *-er* [əʀ] induces resyllabification of the stem-final obstruent as onset of the next syllable, and hence an underlyingly voiced stem-final obstruent remains voiced before *-er* (e.g., Booij, 1995). As a consequence, stems ending in underlyingly voiced obstruents do not have the same segments in isolation as before *-er* (e.g., [ɦɑʀt] - [ɦɑʀdər] 'hard' - 'harder'). We therefore only selected agent nouns and comparatives with stems ending in an underlying voiceless plosive, so that there is no change of the voicing characteristics of the plosive when the stems occur in isolation.

Furthermore, the comparatives and agent nouns in our initial data set occurred with surface frequencies larger than zero. (Token counts in CELEX are based on a corpus of 42.4 million words of written text for Dutch, and on a corpus of 17.9 million words of written and spoken text for English.) From this initial data set of comparatives and agent nouns, we selected those forms that could subsequently be matched to English comparatives or agent nouns that met all the above criteria, and that, in addition, carried the same onset and coda characteristics (simplex versus complex), and that carried the same vowel characteristics (long versus short). The English set of items was used in Part B of this experiment. This selection procedure resulted in a set of 35 Dutch agent nouns and 27 Dutch comparatives (see Appendix A for a list of all Dutch items). Pseudowords were created from these words by changing several phonemes in the stem, while largely respecting the status of onset and coda (simplex versus complex), the vowel length (long versus short), and the restriction that the stem-final consonant is a voiceless plosive [1]. Due to errors, one word (comparative) and one pseudoword eventually had to be removed from the design.

Separate reading lists were created for the comparatives (e.g., *vetter*), the agent nouns (e.g., *werker*), the stems of the comparatives (e.g., *vet*), the stems of the agent nouns (e.g., *werk*), and their pseudoword counterparts. The lists were recorded in a soundproof recording booth by a native male speaker of Dutch, who was

[1] This word-pseudoword matching in our materials was not perfect: We failed to match for the status of the coda for two Dutch items, we failed to match for the status of the onset for one Dutch item, and we failed to match for the length of the vowel for one Dutch item. For one English item, we failed to match for the status of the coda.

naive regarding the purpose of the experiment. Each pseudoword list was read aloud for practice once before recording. The recordings were digitized at 18.9 kHz.

The forms were spliced out of their list using the PRAAT speech-editing software (Boersma & Weenink, 1996). The stems that were produced in isolation functioned as the first type of stimulus in the experiment ('normal' stems, see upper panel of Figure 3.2 for an example). From the complex forms, a second type of stimulus was created: the 'constructed' stems. The constructed stem consisted of the stem of the complex form — in other words, it was the complex form without the suffix *-er* [əʀ]. The point of splicing was located at the onset of the voicing of the schwa following the stem-final consonant. The point of splicing was always located at a zero-crossing. Figure 3.1 shows an example of a complex form (upper panel) and the stem spliced out of that complex form (lower panel).

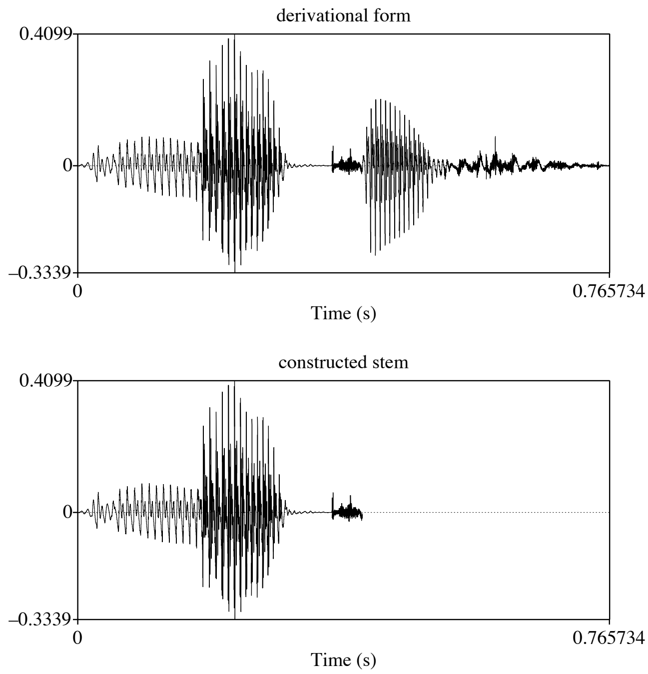

Figure 3.1: The complex form [nɑtəʀ] (upper panel) and the constructed stem [nɑt] spliced out of the complex form (lower panel).

As a result of the splicing manipulation, the constructed stem's prosodic information mismatched its number of syllables: Its prosodic characteristics signalled a bisyllabic form, whereas in fact the acoustic signal contained only one syllable. In the normal stem, there was no such mismatch. Duration was measured for the two types of stems, for both words and pseudowords. As expected, the constructed stems were significantly shorter (161 ms on average) than the normal stems ($F(1,119) = 486.1, p < 0.0001$). The magnitude of this durational difference between normal and constructed stems was not significantly different for words and pseudowords (interaction of Stem Type (normal versus constructed stem) by Word Status (word versus pseudoword): $F(1,119) = 1.6, p = 0.21$). For the words, we also measured the duration of the vowel, the duration of the closure of the stem-final plosive, and the duration of the release noise of the stem-final plosive. Analyses of variance with these durations as the dependent variable, and with Stem Type (normal versus constructed) and the Syllable Structure of the bisyllabic form (with an ambisyllabic stem-final plosive, as in *gok-ker*, 'gambler'; with a syllable-initial stem-final plosive and non-empty coda of the first syllable, as in *hel-per*, 'helper'; with a syllable-initial stem-final plosive and an empty coda of the first syllable, as in *ma-ker*, 'maker') as predictors, revealed significant main effects of Stem Type and Syllable Structure for all three analyses ($p < 0.05$), but never an interaction of these factors ($p > 0.1$). Thus, the manipulation of Stem Type is independent of Syllable Structure.

The normal and constructed stems differed in prosodic structure. The normal and the constructed stems differed in yet another respect, however. The manipulation of interest (the manipulation of prosodic structure) was achieved through and therefore systematically confounded with a splicing manipulation: Splicing had occurred in the constructed stems (at the offset of the release noise of the stem-final consonant), whereas no splicing had occurred in the normal stems. We eliminated this confound by applying a splicing manipulation to the normal stems as well: We spliced away the last 25% of the release noise of the stem-final consonants (see Figure 3.2). As a consequence, both stimulus types ended rather abruptly, the only difference remaining between normal and constructed stems being the difference in prosodic structure. Note that, by applying this splicing manipulation to the normal stems, we put the stimuli that we expected to be most easily processed at a disadvantage. This should make it harder for us to observe an effect of prosodic mismatch. The durational difference between the normal stems and the constructed stems after splicing away 25% of the release noises of the stem-final consonants of the normal stems was 131 ms on average ($F(1,119) = 1391.3, p < 0.0001$). The interaction between Stem Type and Word Status remained non-significant ($F(1,119) = 0.15, p = 0.70$). Table 3.1 lists the mean durations with their standard deviations for the two kinds of stems of words and pseudowords, before as well as after splicing away 25% of the release noise of the normal stems. In the following, the term ‘normal stem’ refers to the stem that carries matching prosodic information *and* of which 25% of the release noise of the stem-final consonant has been spliced away.

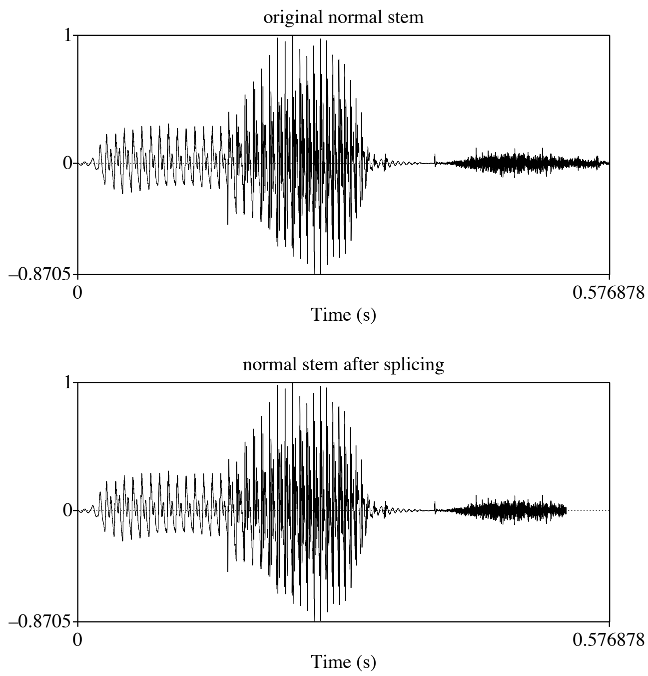

Figure 3.2: The original normal stem [nɑt] (upper panel) and the normal stem [nɑt] after splicing away 25% of the stem-final release noise (lower panel).

Table 3.1: Part A – Mean durations (in ms) with *SD* for normal stems and constructed stems in Dutch, before and after splicing away 25% of the release noise of the normal stems.

Type of stem	Before		After	
	Duration	SD	Duration	SD
Normal word	635	91	597	91
Constructed word	465	79	465	79
Normal pseudoword	593	124	570	97
Constructed pseudoword	441	98	441	98

The total number of experimental trials amounted to 122 (35 agent noun stems and their matched pseudoword stems, and 26 comparative stems and their matched pseudoword stems). So that participants would never be presented with both the normal and the constructed variant of a single stem, complementary versions of trial lists were created. If the normal form of a stem occurred in one version of a list, then the constructed form of that stem would occur in its complementary version. The composition of these lists (i.e., which items occurred in their normal stem variant and which items occurred in their constructed stem variant) was varied three times, resulting in 6 experimental trial lists (three 'compositions' with two complementary versions each). The order of presentation of the stimuli was pseudo-randomized within the three lists: No more than three words or pseudowords occurred successively. Orders were identical in the lists that were each other's complements. Participants were randomly assigned to experimental trial lists. Practice trials were presented prior to the actual experiment. The practice set consisted of 16 trials: 4 normal pseudoword stems, 4 constructed pseudoword stems, 4 normal word stems (2 comparative stems and 2 agent noun stems), and 4 constructed word stems (2 comparative stems and 2 agent noun stems). None of the stems in the practice set was presented in the actual experiment.

Procedure. Participants performed a lexical decision task. They were instructed to decide as quickly as possible whether or not the form that they heard was an existing word of Dutch. They responded by pressing one of two buttons on a button box. Each trial consisted of the presentation of a warning tone (189 Hz) for 500 ms, followed after an interval of 200 ms by the auditory stimulus. Stimuli were presented through Sennheiser headphones. Reaction times were measured from stimulus offset. Each new trial was initiated 2500 ms after offset of the previous stimulus. When a participant did not respond within 2000 ms post-offset, a time-out response was recorded. Prior to the actual experiment, the set of practice trials was presented, followed by a short pause. The total duration of the experimental session was approximately 10 minutes.

Part B: English

Method

Participants. Thirty-nine participants, students at Wayne State University, received course credit to participate in the experiment. All were native speakers of English.

Materials. The selection procedure described above for the Dutch materials resulted in a set of 35 English agent nouns and 27 English comparatives (see Appendix B for a list of all English items). For these words, pseudowords were created by changing several phonemes in the stem, while respecting the status of onset and coda (simplex versus complex), the length of the vowel (long versus short), and the restriction that the stem-final consonant is a voiceless plosive.

Reading lists were created in the same manner as in Part A of the experiment. The lists were recorded in a soundproof recording booth by a native male speaker of English [2]. Each pseudoword list was read aloud for practice once before recording. The recordings were digitized at 20 kHz.

Normal and constructed stems were created in the same manner as in Part A of the experiment. As expected, the constructed stems were again significantly shorter (146 ms) than the normal stems ($F(1, 121) = 937.0, p < 0.0001$). The effect of Stem Type on duration was significantly larger for words than for pseudowords (interaction of Stem Type by Word Status: $F(1, 121) = 7.3, p < 0.01$). Recall that, for Dutch, this interaction of Stem Type by Word Status was not significant, although it did show the same pattern (larger effect of Stem Type for words than for pseudowords). In the overall analysis, the interaction of Stem Type by Word Status was significant ($F(1, 141) = 6.5, p < 0.05$), and there was no significant three-way interaction of Stem Type by Word Status by Language ($F(1, 241) = 0.18, p = 0.67$). We will return to this issue below. Furthermore, the effect of Stem Type on duration was marginally smaller in English than in Dutch (interaction of Stem Type by Language: $F(1, 242) = 3.2, p = 0.07$). As for the Dutch words, we also measured the duration of the vowel, the duration of the closure of the stem-final plosive, and the duration of the release noise of the stem-final plosive for the English words. Analyses of variance with these durations as the dependent variable, and with Stem Type (normal versus constructed) and the Syllable Structure of the bisyllabic form (*cut-ter* versus *hel-per* versus *ma-ker*) as predictors, revealed the following: For the duration of the vowel, there was only a main effect of Stem Type ($p < 0.01$). There was no effect of Syllable Structure nor an interaction of Syllable Structure with Stem Type ($p > 0.1$). None of these factors was predictive for the duration of the release noise. For the duration of the closure, Stem Type was predictive ($p < 0.01$), and there was an interaction of Syllable Structure with Stem Type ($p < 0.01$): For words such as *hel-per*, the difference in closure duration was somewhat less pronounced than for words such as *ma-ker* and *cut-ter*. Thus, the manipulation of Stem Type was independent of Syllable Structure, except for a small difference for one syllable type

[2] In American English, a stem-final /t/ typically becomes flapped in intervocalic position. Our speaker retained the non-flapped pronunciation in intervocalic position, which may be considered overly careful speech. Note, however, that the presence of unflapped stimuli in our experiment should work against our effect, as the unflapped /t/ in the constructed stem might be considered a strong cue for the monosyllabic form.

with respect to closure duration.

The difference in duration between normal and constructed stems remained significant after splicing away 25% of the release noise of the stem-final plosive for the normal stems (121 ms on average; $F(1, 121) = 837.7, p < 0.0001$). Table 3.2 lists the mean durations with their standard deviations for the two kinds of stems of words and pseudowords, before as well as after splicing away 25% of the release noise of the normal stems. The interaction of Stem Type by Word Status was now only marginally significant ($F(1, 242) = 2.9, p = 0.09$), and the three-way interaction of Stem Type, Word Status, and Language remained non-significant ($F(1, 242) = 1.4, p = 0.24$). The effect of Stem Type on duration was still marginally smaller in English than in Dutch (interaction of Stem Type by Language: $F(1, 242) = 2.0, p = 0.09$).

Table 3.2: Part B – Mean durations (in ms) with *SD* for normal stems and constructed stems in English, before and after splicing away 25% of the release noise of the normal stems.

Type of stem	Before		After	
	Duration	SD	Duration	SD
Normal word	506	101	475	97
Constructed word	347	84	347	84
Normal pseudoword	497	91	478	91
Constructed pseudoword	364	89	364	89

Three experimental trial lists and their complements were created in the same manner as in Part A of the experiment. The total number of experimental trials amounted to 124. The practice set consisted of 16 trials: 4 normal pseudoword stems, 4 constructed pseudoword stems, 4 normal word stems (2 comparative stems and 2 agent noun stems), and 4 constructed word stems (2 comparative stems and 2 agent noun stems). None of the stems in the practice set was presented in the actual experiment.

Syllable Ratio and Cohort Entropy were calculated for both the Dutch and the English words. Figures 3.3 and 3.4 summarize the distributions of Syllable Ratio and Cohort Entropy for the agent noun stems and the comparative stems in the Dutch and English part of the experiment, by means of boxplots. Each box shows the interquartile range, the filled circle in the box denotes the median, and the 'whiskers' extend to the observations within 1.5 times the interquartile range. Outliers beyond this range are represented by individual open circles.

Syllable Ratio was significantly higher for English than for Dutch ($F(1, 119) =$

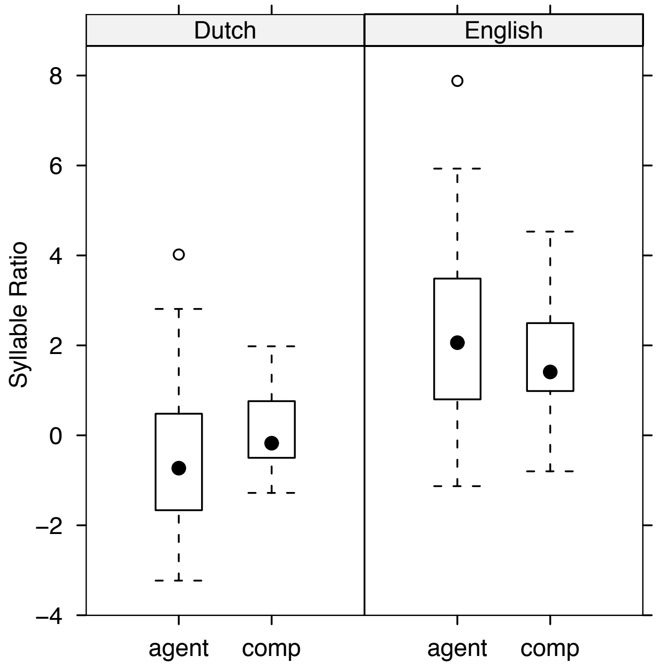

Figure 3.3: Syllable Ratio as a function of Word Type (stem of agent noun versus stem of comparative) and Language (Dutch versus English).

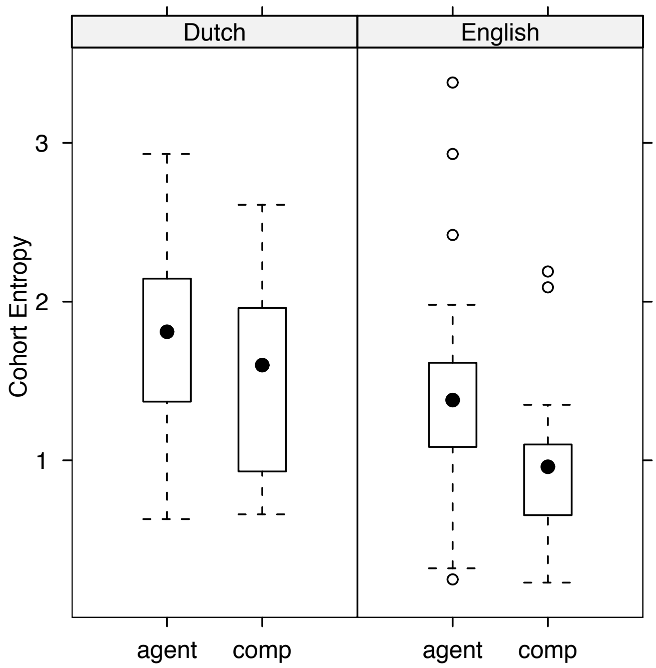

Figure 3.4: Cohort Entropy as a function of Word Type (stem of agent noun versus stem of comparative) and Language (Dutch versus English).

$68.9, p < 0.0001$). This is what we expected, as there are fewer continuation forms with unstressed syllables in English than in Dutch. Word Type (agent noun versus comparative) had a stronger effect in Dutch than in English (with slightly higher Syllable Ratios for comparative stems than for agent noun stems), but this effect failed to reach significance in both languages (interaction of Word Type by Language: $F(1, 119) = 10.1, p < 0.05$; Dutch: $F(1, 59) = 2.1, p = 0.15$; English: $F(1, 60) = 2.2, p = 0.15$). Cohort Entropy was significantly lower for English than for Dutch ($F(1, 119) = 20.2, p < 0.0001$). This was also expected, since there are fewer continuation forms in general in English than in Dutch. Cohort Entropy was significantly lower for comparative stems than for agent noun stems ($F(1, 119) = 11.0, p < 0.01$). This effect of Word Type on Cohort Entropy was similar for English and Dutch (interaction of Word Type by Language: $F(1, 119) = 0.6, p = 0.42$). Furthermore, it turned out that Syllable Ratio and Cohort Entropy were correlated in English (Pearson's $r = -0.24, p = 0.06$), but not in Dutch (Pearson's $r = -0.14, p = 0.29$). Apparently, Cohort Entropy and Syllable Ratio consider largely the same continuation forms in English, but not in Dutch. In English, most continuation forms have unstressed syllables, whereas, in Dutch, many types of continuation forms are possible.

Procedure. Participants performed English lexical decision. The same procedure was followed as in Part A of the experiment.

Results and discussion

For Dutch (Part A), no participants were excluded from the analyses, since they all showed error rates below 20%. Appendix A lists the mean reaction times and the error rates for the Dutch words and pseudowords. Fifteen items (10 existing words and 5 pseudowords) were excluded from subsequent analyses, as they showed error rates above 20%. Of these 15 items, 6 items had high error rates in both stem variants (i.e., normal and constructed), 6 items had high error rates in the normal variant, and 3 items had high error rates in the constructed variant. Furthermore, trials eliciting incorrect responses were excluded (3% of the trials that remained after removal of the 15 items with high error rates), as well as trials eliciting reaction times faster than 150 ms (3% of all remaining correct trials).

For English (Part B), two participants were excluded from the analyses, since they performed with error rates above 20%. Appendix B lists the mean reaction times and the error rates for the English words and pseudowords, calculated over the trials remaining after removal of the two participants with high error rates.

Twenty-five items (8 existing words and 17 pseudowords) were excluded from subsequent analyses, as they showed error rates above 20%. Of these 25 items, 6 items had high error rates in both stem variants, 9 items had high error rates in the normal variant, and 10 items had high error rates in the constructed variant. Finally, trials eliciting incorrect responses (5% of the trials that remained after removal of the two participants and the 25 items with high error rates) and trials eliciting reaction times faster than 150 ms were also excluded (4% of all remaining correct trials).

The mean response latencies (measured from word offset and calculated over the remaining correct trials only), their standard deviations, and the error percentages for the different types of stems for English and Dutch are summarized in Table 3.3. In general, incorrect responses occurred more often for pseudowords than for words ($z = -6.8, p < 0.0001$), and more often for constructed stems than for normal stems ($z = -3.0, p < 0.01$). The effect of Word Status on performance interacted with Language, however ($z = 4.6, p < 0.0001$): It was significant for English ($z = -6.8, p < 0.0001$), but not for Dutch ($z = 1.1, p = 0.29$).

Table 3.3: Mean reaction times from word offset (in ms) with *SD* and error percentages for normal stems and constructed stems in Dutch and English.

Type of stem	Reaction time	SD	Error
Dutch normal word	464	230	6%
Dutch constructed word	515	218	8%
Dutch normal pseudoword	526	238	6%
Dutch constructed pseudoword	596	226	6%
English normal word	335	160	2%
English constructed word	403	184	4%
English normal pseudoword	428	200	7%
English constructed pseudoword	488	215	7%

In the following, we will report on an overall analysis, as well as on analyses of several subsets of the data. We will start with the overall analysis of the dataset including words as well as pseudowords, for Dutch as well as for English. Next, we will report on an analysis of only the pseudoword data for Dutch and English, and on a similar analysis of only the word data for Dutch and English. Finally, we will report on separate analyses for the Dutch and the English word data. The reasons for analyzing each of these different subsets of the data will be clarified as we proceed.

In an initial, overall analysis, the data for Dutch and English words and pseudowords were analyzed together. We fitted a multi-level covariance model (Pinheiro

& Bates, 2000) to the data, with log reaction times [3] as the dependent variable, and Stem Type (normal versus constructed stem), Word Status (word versus pseudoword), Duration (the duration of the form that was actually presented to the participants) [4], and Language (Dutch versus English) as predictors [5]. Note that in this analysis, we used only a subset of the available predictors. Syllable Ratio was not included as a predictor as it is not possible to calculate this ratio for pseudowords. It is possible to calculate Cohort Entropy for both words and pseudowords, but because Cohort Entropy exhibited very different distributions for words and pseudowords, we did not include Cohort Entropy as a predictor in this overall analysis. We will return to this issue below.

This analysis revealed significant effects of all predictors: Constructed stems were responded to slower than normal stems (56 ms on average for Dutch and 64 ms on average for English, $F(1, 5149) = 306.9, p < 0.0001$), pseudowords were responded to slower than words (91 ms on average for Dutch and 89 ms on average for English, $F(1, 5149) = 529.1, p < 0.0001$), duration was facilitatory (the longer the word, the faster the response latencies; $t(5149) = -8.7, p < 0.0001$), and the Dutch participants were slower than the English participants (100 ms on average, $F(1, 55) = 10.8, p < 0.01$). Furthermore, there were significant interactions of Word Status by Language (the effect of Word Status was less strong in Dutch than in English; $F(1, 5149) = 4.2, p < 0.05$), and of Stem Type by Duration (Duration was more facilitatory for the constructed stems; $t(5149) = 2.9, p < 0.01$). To understand the latter interaction, consider that the longer a given constructed stem is, the more

[3]Here and in the following analyses, reaction times were logarithmically transformed in order to normalize their distribution.

[4]As reaction times were measured from word offset, we expect a facilitatory effect of Duration: At word offset, the listener has had more time to process the word when the duration of the word is long than when the duration of the word is short, facilitating the response. In order to establish whether Stem Type has an effect *independently* of Duration (normal stems have longer durations than constructed stems), we included Duration as a covariate in our analyses.

[5]In our multi-level covariance models, subject variability is accounted for by using subject as a grouping factor. In the analyses of word data exclusively, item variability is accounted for by including item-specific covariates in the regression model. However, in all our analyses involving both word and pseudoword data, and in all analyses involving pseudoword data exclusively, item variability has not been accounted for, as no item-specific covariates are available for pseudowords. Therefore, in all analyses involving pseudowords, Stem Type has been treated as a between-items factor even though we would have liked to treat it as a within-items factor. Nevertheless, even without the extra power of the within-items analysis, we obtained very robust effects of Stem Type. Furthermore, an analysis on Dutch and English words and pseudowords with *item* as the grouping factor yielded largely the same pattern of results as the analysis with *subject* as the grouping factor (Stem Type: $F(1, 203) = 117.3, p < 0.0001$; Word Status: $F(1, 203) = 130.2, p < 0.0001$; Duration: $t(203) = -5.2, p < 0.0001$; Language: $F(1, 203) = 138.4, p < 0.0001$; Stem Type by Duration: $t(203) = 2.9, p < 0.01$). The interaction of Word Status by Language was not significant in this analysis ($F(1, 5149) = 4.2, p = 0.67$).

it resembles its normal stem variant. Apparently, the less abnormal a form is, the faster listeners can respond to it.

To conclude, we have replicated the prosodic mismatch effect for stems of agent nouns and comparatives, in both Dutch and English. The prosodic mismatch effect emerged both in words and in pseudowords. Now the question remains: Do Cohort Entropy and Syllable Ratio have any predictive value? This question calls for separate analyses for words and pseudowords, for two reasons. First, Cohort Entropy (calculated at the stem-final segment) turned out to be normally distributed for Dutch and English words, but not for Dutch and English pseudowords: For the majority of pseudoword items, the cohorts were empty at the stem-final segment, and thus, the Cohort Entropy for these items was zero. For only a small number of pseudoword items (14 out of 56 Dutch pseudowords, and 9 out of 45 English pseudowords), the cohort at the stem-final segment was not empty. Second, the predictor Syllable Ratio cannot be calculated for pseudowords.

We first turn to an analysis of the pseudoword data only. Because of the non-normal distribution of Cohort Entropy, we decided to treat Cohort Entropy as a factor with two levels (Entropy Zero versus Entropy Non-Zero), instead of as a co-variate. In a multi-level covariance analysis, log reaction times were analyzed as a linear function of Stem Type (normal versus constructed stem), Cohort Entropy (Entropy Zero versus Entropy Non-Zero), Duration, and Language (Dutch versus English). This analysis revealed significant effects of all predictors: Constructed stems were responded to slower than normal stems ($F(1, 2486) = 152.9, p < 0.0001$), Duration had a facilitatory effect ($t(2486) = -6.5, p < 0.0001$), English reaction times were faster than Dutch reaction times ($F(1, 55) = 7.0, p < 0.05$), and, importantly, items with empty cohorts (Entropy Zero) were responded to faster than items with non-empty cohorts (Entropy Non-Zero; $F(1, 2486) = 41.8, p < 0.0001$). Furthermore, there was a significant interaction of Cohort Entropy with Language: The effect of Cohort Entropy was less strong for English than for Dutch ($F(1, 2486) = 4.4, p < 0.05$). The effect of Cohort Entropy was significant in both languages, however (Dutch: $F(1, 1040) = 36.7, p < 0.0001$; English: $F(1, 1444) = 10.5, p < 0.01$).

We now turn to the word data. Log reaction times to the words were predicted by the same variables as log reaction times to the pseudowords: Stem Type (normal versus constructed stem), Duration, Cohort Entropy, and Language (Dutch versus English). In addition, Word Type (agent noun versus comparative) and Syllable Ratio were introduced as predictors. For the words (as opposed to the pseudowords), the Cohort Entropy values were normally distributed. Therefore, Cohort Entropy was now treated as a covariate (as opposed to as a factor).

A multi-level covariance analysis revealed significant effects of Stem Type (constructed stems were responded to slower than normal stems; $F(1, 2597) = 194.9, p < 0.0001$), Duration (facilitatory effect; $t(2597) = -5.9, p < 0.0001$), Language (English participants were faster than Dutch participants; $F(1, 55) = 11.5, p < 0.01$), and Word Type (adjectives were responded to faster than verb stems; $F(1, 2597) = 9.4, p < 0.01$). Furthermore, there was a significant inhibitory main effect of Cohort Entropy ($t(2597) = 3.1, p < 0.01$), whereas there was no significant main effect of Syllable Ratio ($t(2597) = 1.7, p = 0.08$). In addition, however, there was a significant second-order interaction of Cohort Entropy by Language ($t(2597) = -2.1, p < 0.05$), and a significant third-order interaction of Syllable Ratio by Cohort Entropy by Language ($F(2, 2597) = 9.1, p < 0.0001$). We will return to this issue below. Finally, we observed a significant interaction of Stem Type by Duration: Duration was more facilitatory for the constructed stems ($t(2597) = 3.2, p < 0.01$). This interaction had already been observed in the overall analysis described above (words and pseudowords in Dutch and English): The longer a given constructed stem is, the more it resembles its normal stem variant, and the faster listeners can respond to it.

As mentioned above, Syllable Ratio and Cohort Entropy were correlated in English (Pearson's $r = -0.34, p < 0.05$), but not in Dutch (Pearson's $r = -0.16, p = 0.26$) [6]. This, in combination with the fact that we observed a second-order interaction of Cohort Entropy by Language, and a third-order interaction of Syllable Ratio by Cohort Entropy by Language, calls for separate analyses for the Dutch and the English word data. These separate analyses yielded the following results.

For Dutch, significant effects were again obtained for Stem Type ($F(1, 910) = 10.4, p < 0.01$), for Duration ($t(910) = -3.4, p < 0.001$), and for Word Type ($F(1, 910) = 4.6, p < 0.05$). Syllable Ratio had a significant facilitatory effect ($t(910) = -3.3, p < 0.01$), but there was no effect of Cohort Entropy ($t(910) = 0.1, p = 0.88$). Interestingly, there was a marginally significant interaction of Syllable Ratio by Stem Type ($t(910) = 1.9, p = 0.06$): The effect of Syllable Ratio was highly significant for the constructed stems ($t(458) = -3.5, p < 0.001$), but non-significant for the normal stems ($t(430) = -0.8, p = 0.40$). In other words, listeners only profited from a high Syllable Ratio when the monosyllabic form they were listening to was abnormal. This suggests that when the mapping of the acoustic signal on the representation

[6] The correlation coefficients reported here are calculated over the items that remained after removing the items with high error percentages, and are therefore numerically different from the correlation coefficients reported in the Materials section (which were calculated over all items that were presented to the participants).

of a stem is less effective as a result of the prosodic characteristics of the acoustic signal, the long-term probability of hearing an unshortened stem is more influential than when the bottom-up signal is unambiguous.

For English, a different pattern emerged. We observed the usual effects of Stem Type ($F(1, 1686)$ = $159.8, p < 0.0001$), Duration ($t(1686) = -3.7, p < 0.001$), and Word Type ($F(1, 1686) = 4.7, p < 0.05$). Syllable Ratio, however, did not have a significant effect ($t(1686) = -0.8, p = 0.45$), whereas there was a significant inhibitory effect of Cohort Entropy ($t(1686) = 2.2, p < 0.05$). This effect of Cohort Entropy is an interesting finding, given the fact that the cohorts over which the Cohort Entropy values were calculated consist mainly of morphologically related continuation forms (i.e., inflections, derivations, and compounds). For English, of all 1,488 possible continuation forms (counted over all word stems), 1,280 forms (113,748 tokens) were morphologically related, and 208 forms (12,539 tokens) were morphologically unrelated. Of the morphologically related forms, 990 forms (110,234 tokens) were inflectional or derivational forms, and 290 forms (3,514 tokens) were compounds. In the cohort literature, it is generally assumed that morphological (inflectional and derivational) continuation forms should be excluded from the cohort (e.g., Marslen-Wilson, 1984; Tyler, Marslen-Wilson, Rentoul, & Hanney, 1988). Our finding shows that for a more realistic indication of the amount of competition in the mental lexicon, morphological continuation forms should be counted as cohort members. Unlike Syllable Ratio in Dutch, Cohort Entropy in English did not interact with Stem Type ($t(1685) = 0.17, p = 0.86$).

For completeness, we note that when Cohort Entropy is not included in the model, Syllable Ratio *is* predictive in English ($t(1687) = -2.0, p < 0.05$). However, when both correlated predictors Syllable Ratio and Cohort Entropy are entered into the model, only the latter is significant. In contrast, Cohort Entropy never showed an effect for Dutch, neither in a model with both Cohort Entropy and Syllable Ratio as predictors, nor in a model that included Cohort Entropy but not Syllable Ratio ($t(912) = 0.5, p = 0.61$).

To conclude, Syllable Ratio (a phonologically motivated measure) emerged as the superior predictor for Dutch reaction times, whereas Cohort Entropy (a non-phonologically motivated measure) emerged as the superior predictor for English reaction times. Apparently, in a language in which word stems are frequently followed by unstressed syllables, that is, in which stems frequently occur in shortened form, listeners develop a sensitivity for the likelihood of observing a shortened or an unshortened stem. In a language in which word stems occur relatively infrequently in shortened form, listeners are less sensitive to the likelihood of observing a shortened or an unshortened stem, but are instead sensitive to the contents of the cohort at stem-final position in general.

General discussion

In this study, we replicated the prosodic mismatch effect that was originally observed for plural inflection in Dutch (Kemps et al., submitted; Chapter 2 of this thesis) for another type of inflection (the formation of comparatives) and for derivation (the formation of agent nouns), in both Dutch and English. Listeners were presented with monosyllabic stems of comparatives (adjectives) and monosyllabic stems of agent nouns (verbs) that carried prosodic information that either matched or mismatched the number of syllables: The matching prosodic information pointed to a monosyllabic form, whereas the mismatching prosodic information pointed to a bisyllabic form. Lexical decision latencies were significantly slower for the items with mismatching prosodic information. This prosodic mismatch effect emerged for words as well as for pseudowords.

English is a morphologically less productive language than Dutch. As a consequence, a stem in English occurs less often in shortened form than a stem in Dutch. Nevertheless, our experiments show that Dutch and English listeners are equally sensitive to prosodic cues in the stem that signal whether or not the stem will be followed by one or more unstressed syllables. The difference in morphological richness between Dutch and English is however reflected in the predictive values of Syllable Ratio relative to Cohort Entropy. Dutch listeners are sensitive to Syllable Ratio, the log odds ratio of observing an unshortened form versus observing a shortened form: In the morphologically richer language, listeners are sensitive to the item-specific distribution of shortened and unshortened stems within the lexicon. In the morphologically poorer language, Cohort Entropy (the entropy of the distribution of cohort members at stem-final position) emerged as the superior predictor, and Syllable Ratio did not have any additional predictive value. Apparently, in a language such as English, in which stems occur relatively infrequently in shortened form, listeners are less sensitive to the item-specific distribution of shortened and unshortened stems within the lexicon. Instead, the contents of the (phonologically and morphologically non-restricted) cohort codetermine response latencies.

Our experiments also show that, in Dutch, Syllable Ratio is facilitatory for the constructed stems only. Apparently, when the mapping of the acoustic signal on the representation of a stem is less effective as a result of the prosodic characteristics of the acoustic signal, the long-term probability of hearing an unshortened stem has a larger role to play than when the bottom-up signal is unambiguous.

It might be argued that the prosodic mismatch effect arises purely due to a mismatch with syllable frame information. Consider the situation in which a listener hears the constructed form of *helper* (i.e., *help*). The prosodic cues of the stem might guide the listener to posit a syllable boundary before the stem-final plosive. Assuming that syllable frames are part of the lexical representations of *help* and *hel-per*, the inferred syllable boundary before the *p* in the constructed stem of *helper* would lead to a mismatch with the lexical representation of the stem (*hel-p* mismatches *help*). This line of reasoning predicts that a greater mismatch in syllabic structure should correspond with a greater prosodic mismatch effect. To test this prediction, we considered the three syllable structures exemplified by the words *ma-ker*, *hel-per*, and *cut-ter*. For words of the last type, the mismatch with a potential syllable frame is minimal, since the ambisyllabic stem-final plosive is both stem-final and syllable-final. Hence, the prosodic mismatch effect should be smallest for *cut-ter*, and larger for *ma-ker* and *hel-per* due to the misalignment of morphological and prosodic structure. Analyses of covariance of the response latencies in Dutch and English with Syllable Structure as an additional predictor revealed the following. In Dutch, an interaction of Syllable Structure with Stem Type emerged ($p < 0.05$), indicating that the words with an ambisyllabic stem-final plosive suffered most instead of least from the Stem Type manipulation, contrary to the above prediction. In English, no interaction was present ($p > 0.6$). We conclude that the prosodic mismatch effect cannot be reduced to a syllable frame mismatch effect.

The subsegmental durational effects documented in the present study probably arise during the mapping of the acoustic signal onto the lexicon. It is less clear at what level the effect of Syllable Ratio should be located. One possibility is to assume that it arises post-lexically. In that case, the inflected and derived words containing a given stem as the first constituent would form the sample space over which the (token-frequency based) probability for that stem of being followed by a syllable with a schwa would be estimated. This estimation, which can be conceptualized either as an on-line generalization over stored exemplars (the inflectional and derivational types), or as an implicit generalization represented in the weights of the connections between morphologically related lexical entries, would then take place after the mapping of the acoustic signal onto the lexical entries is completed.

This is a way in which the present results might be incorporated in a model such as Shortlist (Norris, 1994).

To our mind, a post-lexical explanation of the effect of Syllable Ratio has the disadvantage that different aspects of what may well be the same morpho-phonological phenomenon are spread out over different levels of representation and processing. We view the subsegmental durational differences as providing subtle acoustic cues for the probability of a particular syllable structure and for the likelihood of a following phonologically weak suffix. We interpret Syllable Ratio as a complementary frequency-based estimate of the same probabilities. Although it is technically possible to allocate the subsegmental and Syllable Ratio effects to different levels, we feel that this would lead to a generalization being missed.

As an alternative to a post-lexical explanation, we propose that the Syllable Ratio effect can be understood as an intrinsic part of the process mapping the acoustic input onto the lexicon. In this view, the frequency with which the auditory system encounters inflectional and derivational types leaves its traces in the mapping of the acoustic input onto the lexical representations for these types. Such a mapping operation (sensitive to frequency of occurrence as well as subsegmental duration) is conceivabe in an architecture in which lexical representations are associated with phonetically detailed exemplar clouds.

This way of thinking is compatible with the results of Goldinger's study (1998) which suggest that perceptual details of speech are stored in memory and are integral to later perception. In this study, shadowers showed a tendency to spontaneously imitate the acoustic patterns (speakers' voice characteristics) of words and nonwords. Goldinger simulated these data with the strictly episodic MINERVA 2 model (Hintzman, 1986). In this model, which includes a mechanism of random forgetting necessary to avoid an exponential increase in the costs of storage and retrieval, spoken words were represented by vectors of simple elements. Each vector (i.e., each word token) contained 200 elements, of which 50 elements coded details of the speaker's voice that had produced the word. The model correctly predicted the tendency for shadowers to imitate the idiosyncratic acoustic details of speech, and it successfully predicted the response times in the shadowing task. These results strongly suggest the storage of detailed episodes in the mental lexicon. In the Goldinger study, the vector elements coded — among other things — voice characteristics. Vectors with elements coding other acoustic details, like segment durations, fit well within this approach.

Another subsymbolic, exemplar-based model that allows perceptual detail to be stored in memory, is discussed by Johnson (1997). Word-specific prosodic information was implicitly incorporated in a connectionist model. Johnson trained his model on vector-quantized speech data, which contained — among other things — information regarding the durations of the segments. This model correctly anticipated whether the incoming syllable was followed by another (unstressed) syllable or not. The connection weights in this model applied to our data would be higher between relatively *long* stem exemplars and the stem node than between relative *short* stem exemplars and the stem node. In this model, a constructed stem (with relatively short segment durations) would therefore less effectively activate the stem node than a normal stem (with relatively long segment durations). Similarly, more frequently encountered patterns would lead to enhanced performance.

The probabilistic, exemplar-based framework by Pierrehumbert (2001; 2003) offers a symbolic account of the representation of word-specific phonetic detail in the mental lexicon. In this framework, phonetic categories have probability distributions over a parametric phonetic space. These probability distributions consist of memory traces (exemplars), and are gradually built up as speech tokens are encountered and encoded. Word forms, in turn, are viewed as sequences of phonetic categories, and also have probability distributions over temporal sequences of events in the phonetic space: Individual words have exemplar clouds associated with them. Extending this approach, we might imagine that morphologically complex forms will be associated with exemplars with relatively short stem segments, whereas isolated stems will be associated with exemplars with relatively long segments. Constructed stems are further away from the center of the distribution of stem exemplars than normal stems, and will therefore less effectively activate the representation of the stem.

To conclude, the present study provides more evidence for the role of prosodic information in morphological processing: Detailed acoustic information in the stem reveals whether the stem is realized in isolation or as part of a morphologically complex form. In a morphologically rich language like Dutch (compared to English), listeners are in addition sensitive to the likelihood within the morphological paradigm of a word of encountering a specific prosodic manifestation of that word. Although the data that we have presented in the present paper do not allow us to force a choice between different rival theoretical explanations, the most parsimoneous interpretations seem to point to theories in which the mapping of the acoustic input onto the lexical representations is sensitive to both duration and probability of occurrence.

References

Baayen, R. H., Piepenbrock, R., and Van Rijn, H.: 1993, *The CELEX lexical database (CD-ROM)*, Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Boersma, P. and Weenink, D. J. M.: 1996, PRAAT, a system for doing phonetics by computer, version 3.4, *Institute of Phonetic Sciences of the University of Amsterdam, Report* **132**, Available at http://www.fon.hum.uva.nl/praat/.

Booij, G. E.: 1979, Semantic regularities in word formation, *Linguistics* **17**, 985-1001.

Booij, G. E.: 1995, *The phonology of Dutch*, Clarendon Press, Oxford.

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G.: 2002, Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition, *Journal of Experimental Psychology: Human Perception and Performance* **28**, 218–244.

Fowler, C. A.: 1977, *Timing control in speech production*, Indiana University Linguistics Club, Bloomington, Indiana.

Goldinger, S. D.: 1998, Echoes of echoes? An episodic theory of lexical access, *Psychological Review* **105**, 251–279.

Hintzman, D.: 1986, 'Schema extraction' in a multiple-trace memory model, *Psychological Review* **95**, 528–551.

Johnson, K.: 1997, The auditory/perceptual basis for speech segmentation, *Ohio State University Working Papers in Linguistics* **50**, 101–113.

Kemps, R. J. J. K., Ernestus, M., Schreuder, R., and Baayen, R. H.: submitted, Subsegmental and suprasegmental cues for morphological complexity: The case of Dutch plural nouns.

Lehiste, I.: 1972, *Suprasegmentals*, MIT Press, Cambridge, Mass.

Lindblom, B. and Rapp, K.: 1973, Some temporal regularities of spoken Swedish, *PILUS* **21**, 1–59.

Marslen-Wilson, W. D.: 1984, Function and process in spoken word recognition. A tutorial overview, *in* H. Bouma and D. G. Bouwhuis (eds), *Attention and performance X: Control of language processes*, Erlbaum, Hillsdale, pp. 125–150.

Moscoso del Prado Martín, F., Kostić, A., and Baayen, R. H.: in press, Putting the bits together: An information theoretical perspective on morphological processing, *Cognition*.

Nooteboom, S. G.: 1972, *Production and perception of vowel duration: A study of the durational properties of vowels in Dutch*, Unpublished doctoral dissertation, University of Utrecht, Utrecht.

Norris, D.: 1994, Shortlist: A connectionist model of continuous speech recognition, *Cognition* **52**, 189-234.

Pierrehumbert, J.: 2001, Exemplar dynamics: Word frequency, lenition and contrast, *in* J. Bybee and P. Hopper (eds), *Frequency and the emergence of linguistic structure*, John Benjamins Publishing Company, Amsterdam/Philadelphia, pp. 137–157.

Pierrehumbert, J. B.: 2003, Probabilistic phonology: Discrimination and robustness, *in* R. Bod, J. Hay, and S. Jannedy (eds), *Probability theory in linguistics*, The MIT Press, in press, pp. 177-228.

Pinheiro, J. C. and Bates, D. M.: 2000, *Mixed-effects models in S and S-PLUS*, Statistics and Computing. Springer, New York.

Salverda, A., Dahan, D., and McQueen, J.: 2003, The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension, *Cognition* **90**, 51–89.

Shannon, C. E.: 1948, A mathematical theory of communication, *Bell System Technical Journal* **27**, 379-423.

Tyler, L. K., Marslen-Wilson, W., Rentoul, J., and Hanney, P.: 1988, Continuous and discontinuous access in spoken word-recognition: The role of derivational prefixes, *Journal of Memory and Language* **27**, 368–381.

Appendix A

Dutch materials (orthographic representations):

Dutch agent noun stems and matched pseudowords

Word	Cohort Entropy	Syllable Ratio	Normal Stem		Constructed Stem		Pseudoword	Cohort Entropy	Normal Stem		Constructed Stem	
			RT	Error	RT	Error			RT	Error	RT	Error
1. beuk	2.10	-0.99	513	11.11%	489	0.00%	veek	0.43	730	18.18%	662	0.00%
2. bijt	1.95	-0.73	492	27.27%	664	11.11%	peut	2.06	690	11.11%	734	9.09%
3. breek	1.41	-2.66	507	30.00%	695	0.00%	krook	1.75	507	20.00%	562	0.00%
4. denk	1.33	-0.49	437	0.00%	556	10.00%	lart	0.00	578	20.00%	751	30.00%
5. doop	2.46	0.79	692	22.22%	647	54.55%	book	0.70	468	0.00%	560	0.00%
6. dop	1.73	0.64	922	36.36%	639	0.00%	tep	0.00	694	11.11%	638	0.00%
7. drink	1.12	-1.13	578	0.00%	456	0.00%	plink	0.00	487	18.18%	579	0.00%
8. duik	2.07	-1.00	430	0.00%	480	0.00%	ponk	0.00	464	0.00%	615	0.00%
9. dweep	2.49	-2.68	696	44.44%	663	54.55%	sniep	0.00	650	0.00%	591	0.00%
10. eet	2.41	-1.34	514	0.00%	726	11.11%	oot	1.65	642	0.00%	686	0.00%
11. fluit	1.89	-0.62	332	0.00%	447	0.00%	skoot	0.00	411	0.00%	512	0.00%
12. fok	2.31	-1.27	421	0.00%	732	30.00%	guk	0.00	379	10.00%	499	0.00%
13. gok	1.81	-0.37	390	0.00%	558	0.00%	fot	0.00	696	33.33%	734	36.36%
14. haat	1.05	0.75	378	0.00%	457	0.00%	gaak	0.00	467	0.00%	608	0.00%
15. help	0.97	-2.41	418	0.00%	425	0.00%	relt	0.00	671	22.22%	665	9.09%
16. kaart	1.45	2.69	299	0.00%	418	0.00%	peelt	0.00	553	10.00%	723	0.00%
17. kijk	1.31	-0.89	398	0.00%	584	18.18%	liek	2.25	637	0.00%	624	0.00%
18. kraak	2.20	-3.23	396	0.00%	537	11.11%	plook	0.00	472	0.00%	488	0.00%
19. kweek	2.07	-1.96	505	0.00%	534	0.10%	bleet	0.00	607	11.11%	557	0.00%
20. lok	1.81	-2.01	470	20.00%	825	20.00%	taft	0.00	497	10.00%	643	20.00%
21. maak	1.24	-2.85	612	11.11%	653	27.27%	naap	1.32	613	0.00%	599	0.00%
22. melk	1.75	2.81	292	0.00%	387	0.00%	bork	0.00	474	0.00%	551	0.00%
23. muit	1.66	-1.41	772	60.00%	813	90.00%	beep	0.00	712	11.11%	563	0.00%
24. plant	1.92	0.75	374	10.00%	448	0.00%	krint	0.00	642	18.18%	645	11.11%
25. pleit	0.98	-0.44	636	22.22%	652	9.09%	kleip	0.00	504	0.00%	602	0.00%
26. pluk	1.46	-1.92	435	0.00%	665	11.11%	klek	0.00	505	0.00%	572	0.00%
27. rook	1.79	0.36	461	0.00%	416	0.00%	biek	0.73	439	0.00%	634	10.00%
28. schep	2.34	-2.49	365	10.00%	521	0.00%	spik	1.49	727	9.09%	628	11.11%
29. slaap	2.19	-0.14	413	0.00%	421	0.00%	breep	0.00	395	0.00%	590	10.00%
30. spit	2.30	0.22	518	0.00%	694	0.00%	spep	0.00	500	0.00%	486	0.00%
31. sprint	1.47	0.14	366	0.00%	431	0.00%	skrump	0.00	516	0.00%	457	0.00%
32. strip	2.04	2.56	467	0.00%	504	0.00%	strok	0.00	444	0.00%	546	0.00%
33. vent	1.13	4.02	552	11.11%	426	0.00%	benk	0.00	581	9.09%	592	0.00%
34. werk	2.93	0.60	321	0.00%	492	0.00%	birk	0.00	441	10.00%	601	0.00%
35. zet	0.63	-1.41	642	20.00%	576	10.00%	wuk	0.00	541	10.00%	637	10.00%

Dutch comparative stems and matched pseudowords

Word	Cohort Entropy	Syllable Ratio	Normal Stem RT	Normal Stem Error	Constructed Stem RT	Constructed Stem Error	Pseudoword	Cohort Entropy	Normal Stem RT	Normal Stem Error	Constructed Stem RT	Constructed Stem Error
1. bleek	0.73	-0.50	552	0.00%	474	18.18%	vroot	0.00	473	11.11%	592	0.00%
2. bloot	1.77	-0.39	352	9.09%	464	11.11%	snook	0.00	460	0.00%	495	0.00%
3. dik	1.73	-1.00	436	0.00%	605	10.00%	det	0.00	695	0.00%	592	0.00%
4. flink	0.79	0.31	333	0.00%	427	0.00%	vrunt	0.00	428	0.00%	484	0.00%
5. gek	0.96	1.75	404	0.00%	476	0.00%	gip	1.12	458	10.00%	643	0.00%
6. groot	1.67	-0.80	411	9.09%	388	0.00%	glaat	0.00	535	0.00%	739	20.00%
7. heet	1.19	-0.08	475	0.00%	475	0.00%	wuut	0.00	419	0.00%	578	0.00%
8. juist	0.66	1.28	332	0.00%	346	0.00%	paagt	0.00	456	22.22%	407	54.55%
9. kort	1.67	-0.01	419	0.00%	474	9.09%	firt	0.00	414	0.00%	579	0.00%
10. laat	1.62	-1.28	429	0.00%	546	0.00%	voot	0.69	627	0.00%	768	11.11%
11. mat	2.51	-0.82	480	0.00%	578	20.00%	tup	0.00	619	0.00%	685	0.00%
12. nat	2.07	-0.26	281	0.00%	439	0.00%	plik	0.00	579	18.18%	695	0.00%
13. rank	1.58	-1.21	633	22.22%	594	9.09%	wink	2.12	764	22.22%	740	36.36%
14. rijk	2.44	0.91	416	0.00%	458	11.11%	leep	2.14	661	20.00%	584	10.00%
15. rijp	1.99	-0.41	405	0.00%	472	0.00%	tiek	1.06	562	0.00%	796	20.00%
16. scherp	1.50	-0.01	228	0.00%	429	0.00%	stimp	0.00	497	0.00%	519	0.00%
17. sterk	1.33	0.07	521	11.11%	383	0.00%	blask	0.00	410	0.00%	718	11.11%
18. stomp	2.07	-0.41	528	0.00%	487	0.00%	krunt	0.00	445	9.09%	467	0.00%
19. stout	1.80	-0.14	403	0.00%	477	0.00%	praak	0.00	501	0.00%	705	0.00%
20. strikt	0.68	0.76	527	0.00%	657	0.00%	sprent	0.00	434	0.00%	554	0.00%
21. vast	2.61	1.21	488	0.00%	394	0.00%	mork	0.00	397	0.00%	647	0.00%
22. vlak	0.83	1.21	336	0.00%	518	0.00%	blek	0.60	549	0.00%	751	10.00%
23. wit	0.93	-0.89	451	0.00%	537	0.00%	bup	0.00	482	0.00%	643	0.00%
24. zout	1.26	1.98	335	0.00%	427	0.00%	beut	0.00	508	0.00%	647	0.00%
25. zwak	1.96	-0.44	423	0.00%	445	0.00%	slek	0.00	535	10.00%	613	0.00%
26. zwart	0.89	-0.21	362	0.00%	461	0.00%	knesp	0.00	518	0.00%	633	9.09%

Appendix B

English materials (orthographic representations):

English agent noun stems and matched pseudowords

Word	Cohort Entropy	Syllable Ratio	Normal Stem RT	Normal Stem Error	Constructed Stem RT	Constructed Stem Error	Pseudoword	Cohort Entropy	Normal Stem RT	Normal Stem Error	Constructed Stem RT	Constructed Stem Error
1. look	1.28	7.88	298	0.00%	471	0.00%	nop	0.00	488	9.52%	531	25.00%
2. doubt	1.11	1.93	339	6.25%	386	28.57%	doyp	0.00	416	5.26%	419	0.00%
3. smoke	1.43	2.01	249	0.00%	296	0.00%	shneep	0.00	316	0.00%	421	4.76%
4. bank	1.38	2.24	200	0.00%	428	5.56%	gailt	0.00	487	0.00%	495	43.75%
5. reap	2.42	0.90	438	27.78%	493	10.53%	layp	0.00	449	33.33%	481	12.50%
6. hop	1.74	0.67	490	14.29%	397	6.25%	wep	0.87	541	44.44%	572	63.16%
7. drink	1.06	2.06	228	0.00%	426	15.79%	dromp	0.00	425	5.56%	433	36.84%
8. fake	1.12	2.55	288	0.00%	335	0.00%	veet	1.01	376	23.81%	430	6.25%
9. creep	1.54	0.70	315	10.53%	501	0.00%	klope	0.00	584	10.53%	505	16.67%
10. eat	1.19	3.52	281	5.26%	504	11.11%	ope	1.43	602	10.53%	624	11.11%
11. float	1.47	0.40	279	12.50%	471	33.33%	fruke	0.00	464	31.25%	549	4.76%
12. kick	1.48	2.32	321	0.00%	362	0.00%	kak	1.81	393	6.25%	513	9.52%
13. mock	1.98	3.61	370	16.67%	474	15.79%	nep	0.69	492	12.50%	471	9.52%
14. boat	0.81	4.83	331	0.00%	353	0.00%	doot	0.00	431	4.76%	534	6.25%
15. help	1.31	2.58	340	0.00%	402	0.00%	walp	0.00	460	36.84%	555	22.22%
16. sort	0.49	3.76	419	15.79%	513	5.56%	zaylt	0.00	412	0.00%	412	0.00%
17. make	1.00	4.23	409	0.00%	360	0.00%	neek	0.67	497	27.78%	468	15.79%
18. stalk	2.93	0.63	292	6.25%	403	0.00%	stip	1.67	532	4.76%	488	6.25%
19. break	1.60	3.43	239	0.00%	313	0.00%	ploot	0.76	403	0.00%	462	15.79%
20. pack	0.25	0.00	376	10.53%	609	16.67%	tep	0.24	545	16.67%	555	31.58%
21. bake	1.88	-1.13	354	0.00%	414	0.00%	keek	0.00	396	0.00%	514	0.00%
22. milk	0.55	5.93	269	0.00%	351	0.00%	malp	0.00	371	10.53%	523	16.67%
23. beat	1.67	0.24	391	0.00%	497	0.00%	toop	1.27	412	0.00%	550	0.00%
24. plant	1.20	1.33	287	0.00%	333	0.00%	krent	0.00	374	11.76%	465	10.53%
25. skate	0.94	1.75	219	0.00%	383	0.00%	spote	0.00	496	18.75%	505	0.00%
26. track	1.70	3.45	379	5.26%	523	44.44%	brip	0.00	361	6.25%	503	14.29%
27. hike	1.60	0.97	375	0.00%	523	16.67%	hewt	2.37	530	22.22%	583	63.16%
28. snap	1.56	3.44	285	5.56%	383	0.00%	smik	0.00	402	10.53%	487	11.11%
29. sleep	1.17	2.75	301	0.00%	340	12.50%	shrape	0.00	366	0.00%	420	6.25%
30. quit	0.64	3.90	366	0.00%	478	4.76%	kwop	0.00	383	6.25%	469	9.52%
31. sprint	0.32	0.25	296	4.76%	302	0.00%	strent	0.00	433	15.79%	500	11.11%
32. strip	3.38	1.94	392	0.00%	495	0.00%	splik	0.00	438	0.00%	500	26.32%
33. hunt	1.63	-0.41	299	0.00%	353	0.00%	yamp	0.00	370	0.00%	424	4.76%
34. work	1.48	1.39	267	0.00%	330	0.00%	yert	0.00	361	0.00%	453	0.00%
35. cut	1.15	3.97	345	0.00%	329	0.00%	gak	0.00	471	21.05%	488	11.11%

English comparative stems and matched pseudowords

Word	Cohort Entropy	Syllable Ratio	Normal Stem		Constructed Stem		Pseudoword	Cohort Entropy	Normal Stem		Constructed Stem	
			RT	Error	RT	Error			RT	Error	RT	Error
1. bleak	0.54	2.83	441	28.57%	408	25.00%	gloop	0.00	402	5.56%	537	10.53%
2. cute	0.90	1.81	342	0.00%	512	36.84%	goip	0.00	338	12.50%	438	4.76%
3. thick	1.25	1.33	301	0.00%	394	5.56%	thep	0.00	431	5.56%	503	10.53%
4. drunk	1.13	-0.18	279	0.00%	282	0.00%	klunt	0.00	405	4.76%	528	0.00%
5. sick	2.19	1.27	304	0.00%	350	12.50%	zek	0.00	395	0.00%	375	5.56%
6. great	1.06	1.11	256	0.00%	519	4.76%	bluke	0.51	450	4.76%	625	0.00%
7. tight	2.09	0.53	475	0.00%	565	10.53%	powk	0.00	395	12.50%	470	0.00%
8. moist	0.23	1.10	242	0.00%	279	0.00%	newsk	0.00	369	11.11%	457	15.79%
9. dark	1.01	0.87	272	6.25%	386	0.00%	barp	0.00	431	12.50%	548	14.29%
10. late	0.95	-0.80	353	0.00%	446	5.56%	roke	0.00	504	33.33%	662	25.00%
11. fat	0.89	2.37	347	11.11%	572	42.11%	fik	1.73	585	26.32%	583	44.44%
12. wet	0.66	2.75	231	0.00%	400	0.00%	lup	0.00	509	15.79%	529	5.56%
13. pink	0.55	4.53	223	0.00%	319	0.00%	gont	0.00	532	14.29%	502	12.50%
14. weak	1.35	-0.28	348	0.00%	385	0.00%	yate	0.00	474	12.50%	515	14.29%
15. ripe	1.22	0.73	291	0.00%	360	0.00%	loik	0.00	432	0.00%	578	0.00%
16. sharp	1.26	1.32	231	0.00%	356	4.76%	shelk	0.00	549	5.26%	419	5.56%
17. stark	0.61	2.62	505	31.25%	619	38.10%	spelk	0.00	383	9.52%	540	0.00%
18. plump	0.65	1.78	325	4.76%	428	12.50%	krump	0.66	434	27.78%	453	52.63%
19. bright	1.07	1.30	252	0.00%	332	5.56%	kloik	0.00	304	6.25%	451	4.76%
20. strict	1.01	1.65	305	4.76%	403	12.50%	splekt	0.00	376	5.26%	362	0.00%
21. fast	0.79	0.67	175	0.00%	351	0.00%	hesk	0.00	427	5.26%	551	0.00%
22. black	0.96	3.23	241	6.25%	412	14.29%	kret	1.37	322	5.26%	501	0.00%
23. hot	0.39	3.21	242	0.00%	491	6.25%	yek	0.00	377	20.05%	583	0.00%
24. white	0.56	4.06	267	0.00%	428	5.56%	howk	0.00	461	0.00%	464	5.26%
25. quick	1.04	1.41	338	0.00%	410	5.56%	kwep	0.00	379	12.50%	508	0.00%
26. slick	0.71	2.10	384	0.00%	451	0.00%	slek	0.00	409	11.11%	487	15.79%
27. smart	0.98	1.78	193	0.00%	277	0.00%	snalk	0.68	379	6.25%	363	0.00%

The role of segment duration in ambiguity resolution: Singulars and plurals in Dutch

CHAPTER 4

This chapter has been submitted as Rachèl J. J. K. Kemps, Mirjam Ernestus, Robert Schreuder, and R. Harald Baayen: The role of segment duration in ambiguity resolution: Singulars and plurals in Dutch.

Abstract

The present study investigates the production and comprehension of sentences that are segmentally ambiguous as a result of degemination. We had our speaker produce sentences like *Hij spreekt de kerel/kerels soms* ('He sometimes talks to the guy/guys'). Due to degemination of the cluster of *s*-es, the sentences with the noun in the plural form (e.g., *kerels soms*, with the plural noun *kerels*) are segmentally identical to the sentences with the noun in the singular form (e.g., *kerel soms*, with the singular noun *kerel*): [keʀəlsɔms]. However, our measurements on the duration of the [s] provide evidence for 'incomplete degemination': The [s] functioning as plural suffix as well as as onset of the following word (e.g., in *kerels soms*) was longer than the [s] that functioned only as word onset (e.g., in *kerel soms*). Furthermore, we found that there were durational differences between singular forms and the stems of plural forms, and that these differences showed opposite directions in ambiguous and non-ambiguous sentences. In a perception experiment, we show that listeners attend to the duration of the [s] in an attempt to resolve the ambiguity between sequences like *kerel soms* and *kerels soms*. Listeners also show sensitivity to the duration of the stem, but this sensitivity leads to incorrect responses when confronted with ambiguous sentences. A corpus survey showed that such ambiguities arise very infrequently. We conclude that listeners develop a sensitivity to only those subsegmental patterns that are robustly present in the language.

Introduction

Understanding connected speech involves the successful segmentation of the speech stream into individual words. Occasionally, this segmentation into lexical units is complicated by the ambiguity that arises when the string of segments allows for more than one possible segmentation. Such ambiguity can exist temporarily, with following segmental material providing the resolution of the ambiguity. For example, when the phoneme string [kæp], which is ambiguous between the word *cap* and the word *captain* (and some other words), is followed by the phoneme [m], the word *captain* is ruled out as a candidate. In some cases, however, the ambiguity is not resolved by the following segmental input. For example, the Dutch phoneme sequence [keʀəlsɔms] allows for the segmentation *kerel soms* ('guy sometimes', with the singular noun *kerel*) as well as for the segmentation *kerels soms* ('guys sometimes', with the plural noun *kerels*). In this example, the morphological process of plural formation (i.e., the addition of the plural suffix *-s* to the stem) is counteracted by the phonological process of degemination: The cluster of two *s*-es in *kerels soms* is most of the time realized as a single [s]. In Dutch, degemination of clusters of two identical consonants is obligatory within prosodic words, and optional in larger domains (Booij, 1995). When degemination applies, the contrast that is present in the underlying phonological representation between one consonant and a cluster of two identical consonants is neutralized in the phonetic surface form, sometimes leading to persistent ambiguity.

It has been shown, however, that degemination is not an absolute process, but instead a gradual phenomenon. Martens and Quené (1994) carried out a production study in which they had a Dutch speaker produce phrases that formed minimal pairs like *prei scherp - prijs scherp* ('leek sharp' - 'price sharp', [pʀɛisχɛʀp]) at three different speech rates (slow, normal, fast). The members of the minimal pairs were segmentally identical, but differed in their underlying phonological representation (single versus double fricative). This study showed that the degemination of the two fricatives was strongest at the fastest speech rate, but that the deletion of one of the members of the cluster was never complete: Even at the fastest speech rate, a trace of the double fricative (in the form of a longer duration of the fricative) could still be observed in the phonetic surface form.

Incomplete neutralization of phonologically underlying contrasts has been shown for other kinds of phonological processes as well. In many languages, including Catalan, Dutch, German, Polish, and Russian (see, e.g., Kenstowicz, 1994), underlyingly voiced obstruents are devoiced in word-final or syllable-final position.

Several studies have shown that also this process of final devoicing is not absolute. Neutralized obstruents that are underlyingly voiced tend to have more acoustic characteristics of voiced obstruents than neutralized obstruents that are underlyingly voiceless (e.g., Dinnsen & Charles-Luce, 1984; Port & O'Dell, 1985; Slowiaczek & Dinnsen, 1985; Port & Crawford, 1989; Warner, Jongman, Sereno, & Kemps, in press; Ernestus & Baayen, in press).

Similarly, liaison in French leads to incomplete neutralization of underlying phonological contrasts. In French, the general rule is that word-final consonants are not pronounced (although there are many exceptions to this rule). When the next word begins with a vowel, however, the process of liaison applies: The normally latent word-final consonant surfaces, and this consonant undergoes resyllabification as onset of the first syllable of the following word. This sometimes creates ambiguity at the segmental level, as for example in the case of *dernier oignon*. The final consonant [ʁ] of *dernier* is pronounced, and appears in the initial position of the following word. The resulting phoneme sequence is ambiguous between *dernier oignon* ('last onion') and *dernier rognon* ('last kidney'). Spinelli, McQueen, and Cutler (2003) have shown that this ambiguity is not complete: Consonants that have undergone liaison are shorter than word-initial consonants (i.e., [ʁ] is shorter in *dernier oignon* than in *dernier rognon*).

Evidence is accumulating that incomplete neutralization of underlying phonological contrasts is not only observable in production, but that it also plays a role in perception: Listeners are sensitive to the acoustic traces of the underlying phonological contrasts that surface in the phonetic form. Perception studies indicate that listeners, when presented with minimal pairs, can make use of the small durational differences to distinguish underlying voicing characteristics in neutralized positions (Port & O'Dell, 1985; Port & Crawford, 1989; Warner et al., in press). Ernestus and Baayen (in press) have shown that Dutch listeners use incomplete final devoicing as a subphonemic cue for verbal past-tense formation in Dutch, even when there are no compelling reasons for them to do so. In Dutch, the choice of the past-tense allomorph depends on the voicing characteristics of the stem-final obstruent: [tə] is added when the stem underlyingly ends in a voiceless obstruent, whereas [də] is added elsewhere (Booij, 1995). Participants performed a past-tense formation task with pseudo-verbs. Listeners primarily based their responses on lexical analogy (phonological gangs), but there was nevertheless a role for incomplete final devoicing: They more often chose [də] when the final obstruent was realized as slightly voiced than when it was realized as completely voiceless. Finally, Spinelli et al.

(2003) showed listeners' sensitivity to the subphonemic differences between consonants that had undergone liaison (*dernier oignon*) and word-initial consonants (*dernier rognon*). In a cross-modal priming paradigm, French listeners made visual lexical decisions to vowel- or consonant-initial targets (e.g., *oignon*, *rognon*) following both versions of spoken sentences like *C'est le dernier oignon/rognon*. Facilitation was found when the target matched the speaker's intended segmentation, but was weaker when it mismatched the speaker's intended segmentation.

Listeners thus appear to be sensitive to the acoustic correlates of the incomplete neutralization of underlying phonological contrasts. If it is indeed the case that degemination in Dutch is not an absolute process, but instead a gradual one, as argued by Martens and Quené (1994), then it is conceivable that Dutch listeners resolve the ambiguity that arises when degemination applies (e.g., in the case of *prei scherp/prijs scherp* [pʀɛisχɛʀp]) by attending to the subphonemic cues that are indicative of the underlying contrast (i.e., between a single consonant and a cluster of identical consonants).

A first question addressed in the present study was whether we could find evidence in production for 'incomplete degemination' involving a morphologically functional unit, the plural suffix *-s*. We had our speaker produce sentences like *Hij spreekt de kerel/kerels soms* ('He sometimes talks to the guy/guys'). The sequences *kerel soms* (with the singular noun *kerel*) and *kerels soms* (with the plural noun *kerels*) are segmentally identical: [keʀəlsɔms]. The question is whether the [s] in *kerel soms* is acoustically shorter than the [s] in *kerels soms*. If so, this acoustic difference could serve as a valuable cue in perception for the interpretation of number (singular versus plural). In the second part of our study, we investigated whether listeners might indeed make use of this acoustic difference for resolving the ambiguity between *kerel soms* and *kerels soms*: We presented the sentences from the production study to listeners, and asked them to perform a number decision task on the noun in each sentence. If there is indeed 'incomplete degemination', longer durations of the [s] might lead listeners to respond 'plural' rather than 'singular': Listeners might hypothesize that two [s]-es are underlying, and thus that the [s] consists of both the coda (the plural suffix) of the previous word *and* the onset of the following word. Note, however, that listeners' sensitivity to the duration of the [s] is not necessarily an indication of listeners' sensitivity to 'incomplete degemination'. It might simply indicate listeners' sensitivity to the amount of bottom-up evidence for the presence of an [s]: The more evidence there is for the presence of an [s] after the stem, the more likely it is that the stem is followed by the plural suffix, irrespective of whether the following word's initial segment is an [s] or not.

Other acoustic cues that are indicative of the difference in the underlying representation might be present, however. Previous research (Kemps, Ernestus, Schreuder, & Baayen, submitted; Chapter 2 of this thesis) has shown that there are durational and intonational differences between the (segmentally identical) singular form and the stem of the plural form for monosyllabic nouns in Dutch, and that listeners are sensitive to these acoustic differences. This research concerned nouns of which the plural form consists of the stem (i.e., the singular form) plus the plural suffix *-en* [ə(n)] (e.g., *boek - boeken*, 'book' - 'books'). As a result of the addition of an extra syllable to the stem, the stem of the plural form is shorter and has a higher average fundamental frequency than the singular form. In a number decision task, listeners were presented with singular forms and with stems spliced out of plural forms. They were significantly delayed in responding to the stems of the plural forms, because the segmental and prosodic information were in conflict: The segmental information pointed to the singular form, whereas the prosodic (durational and intonational) information pointed to the plural form. A production study by Baayen, McQueen, Dijkstra, and Schreuder (2003) suggests that similar durational differences exist between singular forms and stems of plural forms of nouns that take the plural suffix *-s* [s] (e.g., *appel - appels*, 'apple' - 'apples'), even though the addition of the plural suffix *-s* does not entail the addition of an extra syllable: Again, the stem of the plural form is shorter than the singular form. These prosodic differences between the singular form and the stem of the plural form could possibly serve as another valuable cue for the resolution of the ambiguity between *kerel soms* and *kerels soms*, provided such differences also occur in these ambiguous sequences. This is not self-evident, as the prosodic differences between singulars and stems of plurals observed in the studies described above appear to be the consequence of segmental differences: The stem is shorter when it is followed by a schwa or an [s] than when nothing follows the stem. In the present study, however, the two readings of the ambiguous sentences under investigation were segmentally identical. Hence, it is not clear that there would be systematic prosodic differences between singular forms and the stems of plural forms, and if so, that listeners would be sensitive to such differences.

To summarize, the present study consists of a production part as well as a perception part. In the production part, our speaker's realizations of sentences like *Hij spreekt de kerel/kerels soms* were studied: We investigated whether, in spite of degemination, the [s] consisting of the plural suffix and the following word's initial segment was acoustically longer than the [s] which is only the following word's initial segment. Furthermore, we investigated whether the durational differences between the singular forms and the stems of the plural forms in the ambiguous sentences are comparable to those in the non-ambiguous sentences. In the perception part of the present study, we investigated which cues — if any — listeners use to resolve the ambiguity between *kerel soms* and *kerels soms*: Are listeners sensitive to the acoustic correlates of 'incomplete degemination', to the durational differences between singular forms and the stems of plural forms, or to both?

Experiment 4.1

Method

Participants. Twenty-four participants, mostly students at the University of Nijmegen, were paid to participate in the perception experiment. All were right-handed native speakers of Dutch.

Materials. From the CELEX lexical database (Baayen, Piepenbrock, & Van Rijn, 1993), we selected 42 Dutch nouns that met the following criteria: They were monomorphemic, they had stem frequencies larger than 100, and singular and plural frequencies larger than zero. They consisted of two syllables, the second one of which was unstressed and ended in *-el* [əl], *-em* [əm] or *-er* [əʀ]. Their plurals consisted of the noun stem plus the plural suffix *-s*. Only common-gender nouns were selected. For nouns of common gender, the definite article is *de* for both the singular and the plural form. For nouns of neuter gender, the definite article is *het* for the singular form, and *de* for the plural form. Because participants were to perform a number decision task, and the nouns would be presented in sentences, in singular and plural form, and preceded by the definite article, we selected only common-gender nouns, so that the article would not give the listener information about the number of the following noun. These nouns formed the experimental materials. They are listed in Appendix A.

As filler materials, we selected 84 Dutch nouns that were monomorphemic, and that had stem frequencies, singular frequencies, and plural frequencies in CELEX larger than zero. Their plurals consisted of the noun stem plus the plural suffix *-en* [ə(n)]. Their stems ended in voiceless plosives. The onset and coda of the stem could be simplex or complex, but never empty. The nouns were all common-gender nouns.

As practice materials, we selected 24 Dutch nouns of which 8 met the same criteria as the experimental materials (except that their stems could also end in *-en* [ən]), and of which 16 met the same criteria as the filler materials.

Two reading lists were created containing the experimental nouns and the practice nouns that met the same criteria as the experimental nouns: one list containing the singular forms, and one list containing the plural forms. The nouns were embedded in sentences that all had the same syntactic structure: *personal pronoun - verb - definite article - noun - adverb*. For example, *Hij spreekt de kerel(s) vaak* ('He often talks to the guy(s)'). The adverb's initial segment was a plosive or a fricative, but was never an [s]. The sentence containing the singular form of a noun was always identical to the sentence containing the plural form of that noun. Two additional reading lists were created, that were identical to the lists described above, except for the adverbs, which were replaced by adverbs that had the same number of syllables and the same stress pattern as the original adverbs, but that had [s] as the initial segment. This resulted in ambiguity: Due to degemination, the sentence containing the singular form of a noun was now segmentally identical to the sentence containing the plural form of that noun (e.g., *Hij spreekt de <u>kerel soms</u>* [keʀəlsɔms] versus *Hij spreekt de <u>kerels soms</u>* [keʀəlsɔms]). The orders of sentences within these 4 lists were randomized twice, resulting in 8 reading lists containing experimental nouns.

Two reading lists were created containing the filler nouns and the practice nouns that met the same criteria as the filler nouns: one list containing the singular forms, and one list containing the plural forms. These nouns were embedded in sentences with the syntactic structure: *personal pronoun - verb - definite article - noun - adverbial adjunct*. For example, *Zij kocht de jurk(en) in Amsterdam* ('She bought the dress(es) in Amsterdam'). The adverbial adjunct's initial segment was a vowel or a fricative. This never resulted in ambiguity. Again, the sentence containing the singular form of a noun was identical to the sentence containing the plural form of that noun. The orders of sentences within these two lists were randomized twice, resulting in 4 reading lists containing filler nouns.

All reading lists were recorded in a soundproof recording booth by a native male speaker of Dutch, who was naive regarding the purpose of the experiment. The recordings were digitized at 16 kHz.

For each sentence, the best realization (of two) was selected. For the experimental nouns, the duration of the stem and — where applicable — the duration of the [s] was measured. Table 4.1 lists the mean durations with their standard deviations for singular and plural forms, in both the ambiguous and the non-ambiguous condition.

Table 4.1: Mean duration (in ms) with *SD* of the stem and of the [s] for singular and plural forms, in ambiguous and non-ambiguous context.

Form	Context	Stem	SD Stem	[s]	SD [s]
Singular	Non-ambiguous	359	55	-	-
Plural	Non-ambiguous	333	57	98	18
Singular	Ambiguous	336	54	126	25
Plural	Ambiguous	361	63	169	25

Figures 4.1 and 4.2 summarize the distributions of [s]-Duration and Stem Duration for the ambiguous and the non-ambiguous singulars and plurals, by means of boxplots. Each box shows the interquartile range, the filled circle in the box denotes the median, and the 'whiskers' extend to the observations within 1.5 times the interquartile range. Outliers beyond this range are represented by individual open circles.

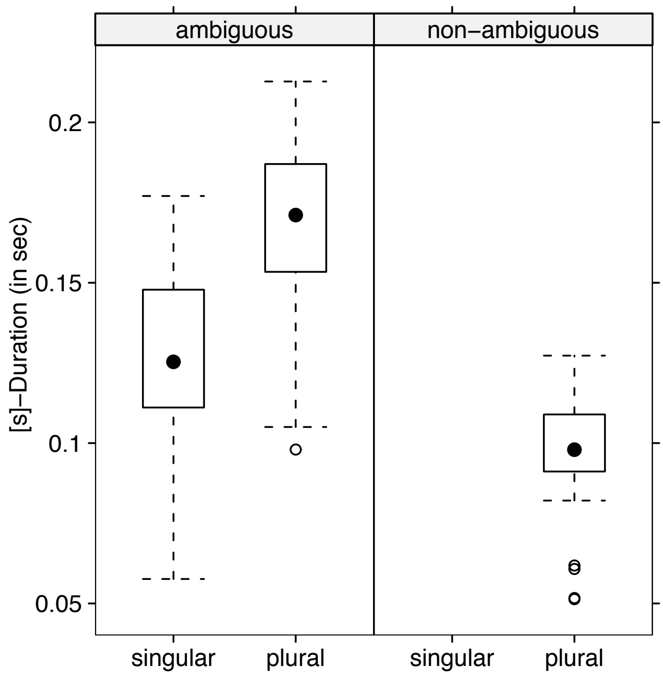

Figure 4.1: [s]-Duration as a function of Ambiguity and Number.

In the ambiguous condition (i.e., *kerel soms* versus *kerels soms*), the duration of the ambiguous [s] was significantly shorter (43 ms on average) in the sentence with the singular form than in the sentence with the plural form ($t(40) = -12.8, p <$

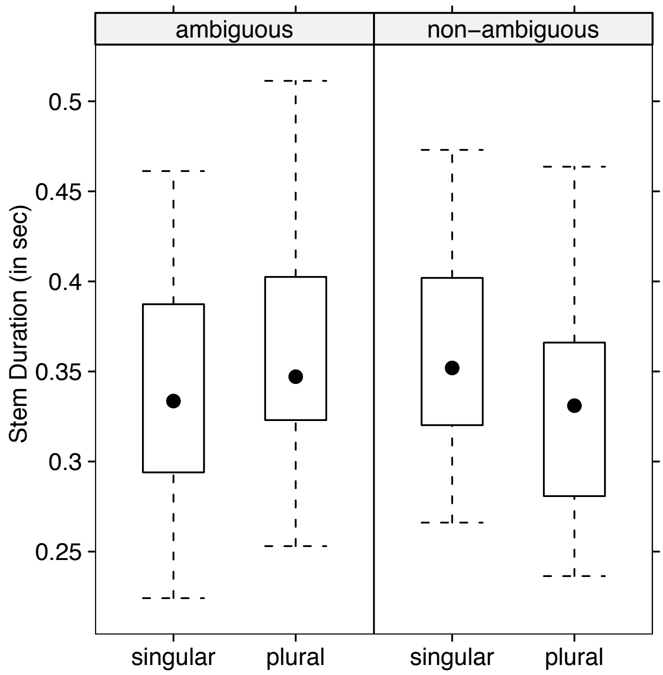

Figure 4.2: Stem Duration as a function of Ambiguity and Number.

0.0001). This shows that the degemination was indeed incomplete [1]: The cluster of two [s]-es was longer than one [s]. The [s] in the plural form in the non-ambiguous condition was significantly shorter (28 ms on average) than the ambiguous [s] in the sentence with the singular form ($t(40) = 6.0, p < 0.0001$).

In the non-ambiguous condition (i.e., *kerel vaak* versus *kerels vaak*), the singular form was significantly longer (27 ms on average) than the stem of the plural form ($t(40) = 7.73, p < 0.0001$). This is consistent with the results by Kemps et al. (submitted; Chapter 2 of this thesis) and by Baayen et al. (2003), mentioned above. Surprisingly, in the ambiguous condition (i.e., *kerel soms* versus *kerels soms*), the reverse was true: The stem of the plural form was significantly longer (25 ms on average) than the singular form ($t(40) = -6.6, p < 0.0001$). This interaction of Number (singular versus plural) by Ambiguity (ambiguous versus non-ambiguous) was highly significant ($F(1, 40) = 100.7, p < 0.0001$).

[1] It might be argued that our speaker did not apply the process of degemination, since this process is optional in domains larger than the prosodic word (Booij, 1995). This seems unlikely, however, as the duration of the cluster of [s]-es in the ambiguous condition (*kerels soms*) was — although longer that the duration of a single [s] in the non-ambiguous condition *kerels vaak* — considerably shorter than the sum of the plural [s] in the non-ambiguous condition and the onset [s] in the singular sentence in the ambiguous condition (*kerel soms*).

For the perception part of this experiment, an experimental trial list was created in such a way that each of the 42 experimental nouns occurred once in each combination of Number and Ambiguity (i.e., singular in non-ambiguous condition, plural in non-ambiguous condition, singular in ambiguous condition, and plural in ambiguous condition). Each of the 84 filler nouns occurred four times, twice in the singular form and twice in the plural form. Each participant was thus presented twice with each filler sentence. This resulted in a trial list consisting of 504 trials, of which 336 trials were filler trials (168 singular forms and 168 plural forms). The other 168 trials were experimental trials, of which 84 trials were ambiguous (42 singular forms and 42 plural forms) and of which 84 trials were non-ambiguous (42 singular forms and 42 plural forms). Thus, 16.7% of all trials (84 of 504) were ambiguous trials. The order of presentation was pseudo-randomized three times (the same noun did not occur in consecutive trials, and ambiguous trials were separated by at least five non-ambiguous or filler trials), resulting in three versions of the experimental trial list.

A practice trial list was created, consisting of the 16 sentences with the practice nouns that take *-en* as the plural suffix and the 8 sentences with the practice nouns that take *-s* as the plural suffix. Half of both types of sentences contained singular forms, half contained plural forms. The order of presentation of the practice trials was randomized three times, resulting in three versions of the practice trial list.

Procedure. Participants performed a number decision task. They were instructed to decide whether the noun in the sentence was in the singular or in the plural form. They responded by pressing one of two buttons on a button box, with the 'singular' button always on the left. As it is conceivable that listeners might attend to different acoustic cues depending on task requirements, half of our participants were instructed to respond as fast as possible, and half of our participants were instructed to respond as accurately as possible. Each trial consisted of the presentation of a warning tone (189 Hz) for 500 ms, followed after an interval of 500 ms by the auditory stimulus. Stimuli were presented through Sennheiser headphones. Reaction times were measured from stimulus offset. Each new trial was initiated 4000 ms after the onset of the previous stimulus. Prior to the actual experiment, the set of practice trials was presented, followed by a short pause. There were three additional pauses in the experiment: a pause of several minutes half-way through the experiment, and two shorter ones after a quarter and three quarters of the experimental trials. The total duration of the experimental session was approximately 40 minutes.

Results and discussion

Due to technical failure, one item had to be excluded from the analyses. In the non-ambiguous condition, 98% of all responses were correct, whereas in the ambiguous condition, 60% of all responses were correct. T-tests on the log odds ratio of correct and incorrect answers showed that participants performed significantly better in the non-ambiguous condition than in the ambiguous condition ($t(60.4) = -19.3, p < 0.0001$), but that even in the ambiguous condition, they performed significantly better than chance ($t(41) = 6.1, p < 0.0001$) [2]. The distribution of correct responses over singular and plural items in the ambiguous and the non-ambiguous condition is presented in Table 4.2. Note that the percentage correct for the ambiguous plurals was 44%. This was marginally significantly worse than chance ($t(41) = -1.8, p = 0.07$). Possibly, some cues in the signal of the ambiguous plurals led participants to respond 'singular', whereas other cues led participants to respond 'plural', leading to correct responses in only 44% of the cases.

Table 4.2: Percentages of correct responses for singular and plural items, in the non-ambiguous and the ambiguous condition.

Condition	Singular	Plural
Non-ambiguous	99%	96%
Ambiguous	76%	44%

The responses were further analyzed using step-wise logistic regression. We were particularly interested in the effect of three factors on the response behaviour of our partipants when presented with sequences that were segmentally ambiguous between singular and plural.

The first factor that we were interested in is the duration of the [s]: Does a long [s] guide listeners to respond 'plural' rather than 'singular'? Recall that, if it does, there

[2]Note that it is impossible to determine on the basis of the *t*-test reported here on the difference in performance between the ambiguous and the non-ambiguous condition whether this difference is the result of a true difference in sensitivity (i.e., singulars and plurals are easier to tell apart in the non-ambiguous than in the ambiguous condition) or instead of a difference in response bias. Similarly, the *t*-test on the responses to the singulars and the plurals in the ambiguous condition does not unequivocally show that the above-chance performance in this condition really reflects listeners' ability to tell the singulars and the plurals apart in this condition. However, if only response bias were at play in the ambiguous condition, the overall performance in that condition would be at 50% and, furthermore, there would not be an effect of [s]-duration in the regression analyses. This is contrary to fact.

would be two possible explanations. One would be that listeners are in fact sensitive to 'incomplete degemination'. When confronted with a rather long [s], listeners might hypothesize that two [s]-es are underlying, and thus that the [s] consists of both the coda (the plural suffix) of the previous word *and* the onset of the following word. Another explanation is possible, however. Below, we will present corpus-based counts of the instances in which the stem is followed by an [s], showing that the instances in which the [s] is the onset of the following word are relatively rare compared to the instances in which the [s] is the plural suffix. A longer [s] means that there is more bottom-up evidence for the presence of an [s], and since the presence of an [s] usually indicates that the form is a plural form, listeners might more often opt for 'plural' with longer durations of the [s].

The second factor that we were interested in is the duration of the stem. Note that — although stem duration has proven to be a valuable cue for number in previous experiments (Kemps et al., submitted; Chapter 2 of this thesis) — the duration of the stem has little local cue validity *within* this experiment, as a long stem duration corresponded to the singular form only half of the time (in the non-ambiguous sentences), and to the plural form the other half of the time (due to the unexpected reversal in the direction of the durational difference between singulars and plurals in the ambiguous sentences). It is therefore conceivable that stem duration will not serve as a valuable cue within this experiment. If it does, however, it would be interesting to see whether it would interact with ambiguity.

Finally, we were interested in what the influence would be of the long-term probability that a given stem is in the singular or in the plural form. Frequency Ratio (the log ratio of the singular and plural frequencies) was therefore introduced as a predictor. We preferred including Frequency Ratio as a predictor over including singular frequency and plural frequency separately for two reasons. First, including singular frequency and plural frequency separately would lead to collinearity in the regression. Second, Frequency Ratio is an index of the frequency of the one form *relative* to the other form. As the task at hand is a number decision task, this relative frequency measure is presumably more relevant than the absolute singular or plural frequency.

In an overall analysis (including ambiguous and non-ambiguous singulars and plurals), the logit of the numbers of correct and incorrect answers was modelled as a linear function of Ambiguity (ambiguous versus non-ambiguous), Number (singular versus plural), Instruction ('fast' versus 'accurate'), Stem Duration, and Frequency Ratio. This analysis yielded significant main effects of Ambiguity (performance was better in the non-ambiguous condition; $Z = 5.6, p < 0.0001$), Instruction (performance was better when participants were instructed to respond accurately; $Z = -2.6, p < 0.01$), and Stem Duration (performance was better with longer Stem Duration; $Z = 3.1, p < 0.01$). Number did not have a significant main effect, but it did emerge as a predictor in interactions with Stem Duration (Stem Duration was facilitatory for singular forms only; $Z = -2.6, p < 0.01$), with Instruction (receiving a 'fast' instruction led to worse performance for the singular forms only; $Z = 2.6, p < 0.01$), and with Frequency Ratio (a high Frequency Ratio led to worse performance for the plural forms only; $Z = -4.2, p < 0.0001$).

In addition, we ran a separate analysis for the ambiguous items, as [s]-Duration could be included as a predictor for the ambiguous items only (non-ambiguous singulars do not contain an [s]). This analysis yielded significant main effects of Number (performance was better for singulars than for plurals; $Z = -4.5, p < 0.0001$), Instruction (performance was better when participants were instructed to respond accurately than when they were instructed to respond fast; $Z = -2.8, p < 0.01$), Stem Duration (performance was better with longer Stem Duration; $Z = 2.1, p < 0.05$), [s]-Duration (performance was better with shorter [s]-Duration; $Z = -3.3, p < 0.001$), and Frequency Ratio (performance was better with a higher Frequency Ratio; $Z = 2.4, p < 0.05$). Furthermore, there were significant interactions of Number by Instruction (being instructed to respond fast led to worse performance for the singular forms only; $Z = 2.8, p < 0.01$), of Number by [s]-Duration (long [s]-Duration led to worse performance for singulars but to better performance for plurals; $Z = 6.1, p < 0.0001$), and of Number by Frequency Ratio (a high Frequency Ratio led to better performance for singulars but to worse performance for plurals; $Z = -4.8, p < 0.0001$).

In order to make the pattern in the data more accessible, the results described above are summarized in Table 4.3. The results are readily interpretable. The main effect of Number that emerged in the sub-analysis of the ambiguous items suggests that there was a bias for responding 'singular': Listeners more often incorrectly classified the plural form as a singular than that they incorrectly classified the singular form as a plural. This bias is possibly a result of the fact that, in general, the singular form is more frequent than the plural form (making the plural form the marked form).

In addition, both analyses show that, overall, listeners performed better when they were instructed to respond as accurately as possible than when they were instructed to respond as fast as possible. Both analyses suggest that this effect of

Table 4.3: Summary of the results of the analyses of the response data (n.s. = non-significant).

Independent Variable	Overall	Ambiguous items
Ambiguity	$p < 0.0001$	not applicable
Number	n.s.	$p < 0.0001$
Instruction	$p < 0.01$	$p < 0.01$
[s]-Duration	not applicable	$p < 0.01$
Stem Duration	$p < 0.01$	$p < 0.05$
Frequency Ratio	n.s.	$p < 0.05$
Number x Instruction	$p < 0.01$	$p < 0.01$
Number x Stem Duration	$p < 0.01$	n.s.
Number x Frequency Ratio	$p < 0.0001$	$p < 0.0001$
Number x [s]-Duration	not applicable	$p < 0.0001$

Instruction was carried mainly by the singular forms: Performance was better for the singulars than for the plurals, and having to respond quickly worsened performance more for the singulars than it did for the plurals.

Furthermore, we observed that a high Frequency Ratio led to better performance for singulars but to worse performance for plurals. In other words, the more the frequency of the singular form exceeds that of the plural form, the easier it is to correctly classify the singular form. The less the frequency of the singular form exceeds that of the plural form, or when the plural frequency is larger than the singular frequency, the easier it is to correctly classify the plural form.

In the analysis of the ambiguous items, we observed that performance was better with longer [s]-Duration for the plurals but worse for the singulars. This is as expected. A longer [s]-Duration means more bottom-up evidence for the presence of an [s] and an [s] following the stem is usually the plural suffix. Furthermore, a longer [s]-Duration means that it is more likely that two [s]-es are underlying, and thus that it is more likely that the form is a plural form. Therefore, a singular form with a relatively long [s] will more often be incorrectly classified as a plural, whereas a plural form with a relatively short [s] will more often be incorrectly classified as a singular.

Finally, we observed an effect of Stem Duration and an interaction of Stem Duration by Number in the overall analysis. Longer Stem Duration led to better performance for the singular forms only: The longer the Stem Duration, the easier it is to correctly classify the singular form. This is consistent with the fact that Stem Duration was on average longer for the singular forms than for the plural forms in the non-ambiguous condition. Recall, however, that the opposite durational pattern was observed in the ambiguous condition: Stem Duration was on average *shorter* for the singular forms than for the plural forms. Nevertheless, as indicated by the absence of a third-order interaction of Stem Duration by Number by Ambiguity, listeners interpreted Stem Duration similarly in the ambiguous and non-ambiguous sentences. For both the ambiguous and the non-ambiguous singulars, longer stem duration led to more correct (i.e., ‘singular’) responses. This may seem surprising, but — as we shall see below — ambiguous sequences like the ones studied here are relatively rare in the language. The durational distribution in the non-ambiguous condition is more representative of the pattern present in the language. Our data suggest that listeners have (implicit) knowledge of the latter, representative distribution of stem durations in the language, and no (implicit) knowledge of the reversal in duration in the ambiguous condition.

We now turn to the analyses of the reaction times. Only correct responses were analyzed, and ‘singular’ and ‘plural’ responses were analyzed separately, because executing these responses involved different hands. The results of these analyses are summarized in Table 4.4.

Table 4.4: Summary of the results of the analyses of the reaction time data (n.s. = non-significant).

Independent Variable	‘Singular’ responses	‘Plural’ responses
Ambiguity	$p < 0.0001$	n.s.
Instruction	$p < 0.0001$	$p < 0.001$
[s]-Duration	not applicable	$p < 0.05$
Stem Duration	n.s.	n.s.
Frequency Ratio	$p < 0.0001$	n.s.
Ambiguity x Instruction	$p < 0.0001$	$p < 0.01$
Ambiguity x Frequency Ratio	$p < 0.01$	n.s.
Instruction x Frequency Ratio	n.s.	$p < 0.0001$

In a multi-level covariance analysis (Pinheiro & Bates, 2000), log reaction times for the ‘singular’ responses (as measured from word offset) were analyzed as a linear function of Ambiguity (ambiguous versus non-ambiguous), Instruction (‘fast’ versus ‘accurate’), Stem Duration, and Frequency Ratio. This analysis revealed significant main effects of Ambiguity (responses were faster in the non-ambiguous condition; $F(1, 1692) = -13.9, p < 0.0001$), Instruction (responses were faster when participants were instructed to respond fast; $F(1, 22) = 27.3, p < 0.0001$), and Frequency Ratio (responses were faster with higher Frequency Ratio; $t(1691) = -5.1, p < 0.0001$). Furthermore, there were significant interactions of Ambiguity by Instruction (the effect of Ambiguity was larger when participants were instructed to respond fast; $F(1, 1691) = 25.0, p < 0.0001$), and of Ambiguity by Frequency Ratio

(Frequency Ratio was more facilitatory in the ambiguous condition than in the non-ambiguous condition; $t(1691) = 3.2, p < 0.01$). The effect of Stem Duration was in the expected direction given the response data (faster responses with longer Stem Duration), but this effect did not reach significance ($t(1691) = -1.2, p = 0.23$).

A similar analysis was run for the 'plural' responses. Log reaction times for the 'plural' responses (as measured from [s]-offset — note that, for the ambiguous items, this [s] is not only the final segment of the plural, but also the onset of the following word) were analyzed as a linear function of Ambiguity (ambiguous versus non-ambiguous), Instruction ('fast' versus 'accurate'), Stem Duration, [s]-Duration, and Frequency Ratio. This analysis yielded significant main effects of Instruction (responses were faster when participants were instructed to respond fast; $F(1, 22) = 19.3, p < 0.001$), and of [s]-Duration (responses were faster with longer [s]-Duration; $t(1341) = -2.1, p < 0.05$). In addition, there was a significant interaction of Instruction by Ambiguity ($F(2, 1341) = 13.0, p < 0.0001$). When participants were instructed to respond fast, responses were slower in the non-ambiguous condition than in the ambiguous condition ($t(1341) = 2.7, p < 0.01$). When they were instructed to respond accurately, there was no effect of Ambiguity ($t(1341) = -1.6, p = 0.12$). Note that the direction of this interaction is opposite to the direction of the interaction of Ambiguity by Instruction that was observed for the 'singular' responses: For the 'singular' responses, when participants were instructed to respond fast, the responses to the non-ambiguous items were faster than those to the ambiguous items. A possible explanation is that the duration of the [s] in the non-ambiguous plural forms is short relative to the duration of the [s] in the ambiguous plural forms. Since [s]-Duration has a significant facilitatory effect for plurals, the effect of Ambiguity we observe here might actually mirror the effect of the [s]-Duration. Finally, there was an interaction of Instruction by Frequency Ratio ($F(1, 1341) = 10.7, p < 0.0001$): Frequency Ratio inhibited participants when they were instructed to respond fast ($t(1341) = 4.4, p < 0.0001$), but not when they were instructed to respond accurately ($t(1341) = 1.3, p = 0.19$). There was no significant effect of Stem Duration ($t(1340) = -1.7, p = 0.09$).

The results of the analyses of the reaction times are generally in line with the results of the analyses of the response data. Performance was better and reaction times were longer when participants were instructed to respond accurately than when they were instructed to respond fast. A high Frequency Ratio led to more correct and faster responses to the singular forms, but to more incorrect and slower responses to the plural forms (at least when participants were instructed to respond fast): The more frequent the singular form is relative to the plural form, the easier (in terms of performance as well as in terms of reaction times) it is to correctly classify the singular form, and the more difficult it is to correctly classify the plural form. A long [s]-Duration led to more incorrect responses for the ambiguous singular forms, to more correct responses for the ambiguous plural forms, and to shorter reaction times for the correct 'plural' responses in both the ambiguous and the non-ambiguous sentences.

To conclude, our listeners showed sensitivity to the duration of the segment [s], and to the durational differences between singular forms and the stems of plural forms. They were, however, not capable of tuning these sensitivities to the specific durational patterns present in the experiment. For optimal performance, our listeners should have interpreted the combination of a long [s] and a long duration of the stem as indicating the plural form, but this is not what they did. Instead, they always interpreted a long stem duration as evidence for the singular form, and were thus faced with conflicting cues in the ambiguous plural sentences: The long stem duration was interpreted as evidence for the singular form, whereas the long duration of the [s] was interpreted as evidence for the plural form.

An estimation of the prevalence in the Dutch language of ambiguous sequences like the ones studied here, based on a corpus of newspaper issues (from the Dutch newspaper 'Trouw', corpus size approximately 5 million words), suggests why listeners are not sensitive to the specific, reversed durational pattern in the noun stems in the ambiguous sentences. We counted the number of times a sequence consisting of a noun ending in *-el* [əl], *-em* [əm] or *-er* [əʀ] followed by another word was one of the following types: non-ambiguous singular (*kerel vaak*), non-ambiguous plural (*kerels vaak*), ambiguous singular (*kerel soms*), ambiguous plural (*kerels soms*). The counts were as follows: 417 non-ambiguous singular cases, 202 non-ambiguous plural cases, 12 ambiguous singular cases, and 4 ambiguous plural cases. This shows, as mentioned above, that the ambiguous cases are relatively rare compared to the non-ambiguous cases. Apparently, the durational pattern associated with the ambiguous singulars and plurals is too infrequent to develop a robust representation in the mental lexicon.

Summary and conclusions

The present study consisted of a production part as well as a perception part. In the production part, we investigated whether sentences like *Hij spreekt de kerel/kerels* *soms*, that are ambiguous at the segmental level, are also ambiguous at the subsegmental level. Our study provides additional evidence that the phonological process of degemination in Dutch is gradual instead of absolute (see also Martens & Quené, 1994). If degemination had been complete, the duration of the [s] should not have differed between the two types of sequences. In fact, the [s] was significantly longer in sequences like *kerels soms* than in sequences like *kerel soms*. Furthermore, in the non-ambiguous sentences, we observed the expected durational difference between singular forms and the stems of plural forms: The stems of the plural forms were shorter than the singular forms. This is consistent with the durational pattern observed by Baayen et al. (2003) and by Kemps et al. (submitted; Chapter 2 of this thesis). The addition of the plural suffix [s] thus shortens the noun stem, also when the noun is realized in sentence context. Surprisingly, in the ambiguous sentences, we observed the reverse pattern: The stems of the plural forms were *longer* than the singular forms.

In the perception part of our study, we investigated whether listeners are sensitive to these acoustic differences between singulars and plurals, and between ambiguous and non-ambiguous sequences. Which acoustic cues — if any — do listeners use to resolve the ambiguity between *kerel soms* and *kerels soms*?

Listeners performed better than chance in the ambiguous condition, which suggests that the sentences in the ambiguous condition were not completely ambiguous to the listeners: The acoustic signal appears to have contained useful cues regarding the number of the noun. The duration of the [s] (or acoustic correlates of the duration of the [s]) appears to be one of these cues. Performance was better (and faster) with long [s]-Duration for plurals, but performance was worse with long [s]-Duration for singulars. This may indicate either listeners' sensitivity to 'incomplete degemination' or simply listeners' sensitivity to the amount of bottom-up evidence for the presence of an [s]. We cannot rule out either explanation based on our results.

Another acoustic cue that listeners used to choose between singular and plural is the duration of the stem (or acoustic correlates of the duration of the stem). However, listeners do not seem to have taken the difference in the durational pattern between non-ambiguous sentences and ambiguous sentences into account: They tended to respond 'singular' when they heard a relatively long stem, irrespective of whether this stem was embedded in an ambiguous sentence or not. This explains why the performance for the singular forms in the ambiguous condition was significantly above chance (76% correct), whereas performance for the plural forms in the ambiguous condition was marginally significantly below chance (44% correct). There were more factors promoting 'singular' responses to the singular forms (general response bias, [s]-Duration, but not stem duration) than there were factors promoting 'plural' responses to the plural forms ([s]-Duration, but neither stem duration nor general response bias).

Listeners thus do not appear to be sensitive to the specific durational patterns in the stem that our speaker produced. Possibly, our speaker was atypical, and produced durational patterns that diverge from what the average speaker would have produced. However, we can think of no reasons why our speaker would have behaved atypically. There is another, more plausible explanation. The corpus-based counts presented above show that ambiguous sentences of the type discussed in this study are very infrequent in the Dutch language. This suggests that the durational distributions specific for singulars and stems of plurals in ambiguous sentences are too infrequent to leave traces in the mental lexicon. It is only the more frequent patterns that are robust enough in the language to be picked up.

To conclude, our listeners attempted to resolve the ambiguity between *kerel soms* and *kerels soms* by attending to the duration of the [s]. They were also sensitive to the robust difference in stem duration between singulars and plurals. These findings are consistent with previous studies showing listeners' capability to pick up very subtle subsegmental cues in the speech signal (e.g., Davis, Marslen-Wilson, & Gaskell, 2002; Salverda, Dahan, & McQueen, 2003; Warner et al., in press; Ernestus, & Baayen, in press; Spinelli et al., 2003; Kemps et al., submitted — Chapter 2 of this thesis; Kemps, Wurm, Ernestus, Schreuder, & Baayen, in press — Chapter 3 of this thesis). This capability is striking, given the enormous variability in the temporal structure of speech. The present study, however, also shows that there are limits to the patterns of cues that listeners can pick up. The sensitivity to the duration of the stem, for example, has an adverse effect on the comprehension of the ambiguous sentences in our experiment, as the difference in stem duration between singulars and plurals in these ambiguous sentences goes in the opposite direction of the default pattern. Listeners are not sensitive to unusual and seldom encountered patterns. They seem to develop a sensitivity to only those patterns of subsegmental cues that are robustly present in the language.

References

Baayen, R. H., McQueen, J., Dijkstra, T., and Schreuder, R.: 2003, Frequency effects in regular inflectional morphology: Revisiting Dutch plurals, *in* R. H. Baayen and R. Schreuder (eds), *Morphological structure in language processing*, Mouton de Gruyter, Berlin, pp. 355–390.

Baayen, R. H., Piepenbrock, R., and Van Rijn, H.: 1993, *The CELEX lexical database (CD-ROM)*, Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Booij, G. E.: 1995, *The phonology of Dutch*, Clarendon Press, Oxford.

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G.: 2002, Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition, *Journal of Experimental Psychology: Human Perception and Performance* **28**, 218–244.

Dinnsen, D. A. and Charles-Luce, J.: 1984, Phonological neutralization, phonetic implementation and individual differences, *Journal of Phonetics* **12**, 49–60.

Ernestus, M. and Baayen, H.: in press, The functionality of incomplete neutralization in Dutch: The case of past-tense formation, *Papers in Laboratory Phonology VIII*.

Kemps, R. J. J. K., Ernestus, M., Schreuder, R., and Baayen, R. H.: submitted, Prosodic cues for morphological complexity: The case of Dutch plural nouns.

Kemps, R. J. J. K., Wurm, L. H., Ernestus, M., Schreuder, R., and Baayen, R. H.: in press, Prosodic cues for morphological complexity in Dutch and English, *Language and Cognitive Processes*.

Kenstowicz, M.: 1994, *Phonology in generative grammar*, Blackwell, Cambridge MA & Oxford UK.

Martens, L. and Quené, H.: 1994, Degemination of Dutch fricatives in three different speech rates, *in* R. Bok-Bennema and C. Cremers (eds), *Linguistics in the Netherlands 1994*, John Benjamins Publishing Company, Amsterdam.

Pinheiro, J. C. and Bates, D. M.: 2000, *Mixed-effects models in S and S-PLUS*, Statistics and Computing, Springer, New York.

Port, R. and Crawford, P.: 1989, Incomplete neutralization and pragmatics in German, *Journal of Phonetics* **17**, 257–282.

Port, R. F. and O'Dell, M. L.: 1985, Neutralization of syllable-final voicing in German, *Journal of Phonetics* **13**, 455–471.

Salverda, A., Dahan, D., and McQueen, J.: 2003, The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension, *Cognition* **90**, 51–89.

Slowiaczek, L. and Dinnsen, D. A.: 1985, On the neutralizing status of Polish word-final devoicing, *Journal of Phonetics* **13**, 325–341.

Spinelli, E., McQueen, J., and Cutler, A.: 2003, Processing resyllabified words in French, *Journal of Memory and Language* **48**, 233–254.

Warner, N., Jongman, A., Sereno, J., and Kemps, R.: in press, Incomplete neutralization and other subphonemic durational differences in production and perception, *Journal of Phonetics*.

Appendix A

The experimental materials (orthographic representations):

Singular	Plural	Singular	Plural
1. anjer	anjers	22. navel	navels
2. beitel	beitels	23. ober	obers
3. bezem	bezems	24. oksel	oksels
4. bliksem	bliksems	25. panter	panters
5. bloesem	bloesems	26. parel	parels
6. bochel	bochels	27. pater	paters
7. bodem	bodems	28. puzzel	puzzels
8. borrel	borrels	29. ridder	ridders
9. drempel	drempels	30. schilder	schilders
10. egel	egels	31. sikkel	sikkels
11. ekster	eksters	32. steiger	steigers
12. ezel	ezels	33. stengel	stengels
13. hamer	hamers	34. tegel	tegels
14. hengel	hengels	35. tunnel	tunnels
15. karper	karpers	36. veter	veters
16. keizer	keizers	37. vinger	vingers
17. kerel	kerels	38. wezel	wezels
18. kikker	kikkers	39. winkel	winkels
19. koker	kokers	40. wortel	wortels
20. koster	kosters	41. zolder	zolders
21. moeder	moeders	42. zuster	zusters

Processing reduced word forms: The suffix restoration effect

CHAPTER 5

This chapter is a revised version of a paper that has been published as Rachèl Kemps, Mirjam Ernestus, Robert Schreuder, and Harald Baayen (2004): Processing reduced word forms: The suffix restoration effect, *Brain and Language*, **19**, 117-127.

Abstract

Listeners cannot recognize highly reduced word forms in isolation, but they can do so when these forms are presented in context (Ernestus, Baayen, & Schreuder, 2002). This suggests that not all possible surface forms of words have equal status in the mental lexicon. The present study shows that the reduced forms are linked to the canonical representations in the mental lexicon, and that these latter representations induce reconstruction processes. Listeners restore suffixes that are partly or completely missing in reduced word forms. A series of phoneme-monitoring experiments reveals the nature of this restoration: The basis for suffix restoration is mainly phonological in nature.

Introduction

In spontaneous speech, words are often produced with fewer segments, or even syllables, than when they are carefully pronounced in isolation. Highly frequent words in particular may occur in highly reduced form. For instance, the Dutch word *eigenlijk* ('actually'), with the canonical pronunciation [ɛiχələk], may in casual speech be realized as [ɛiχək], and the Dutch word *natuurlijk* ('of course'), with the canonical pronunciation [natyʀlək], may be realized as [tyk] (Ernestus, 2000). In both examples, the suffix *-lijk* [lək] is reduced to [k]. Ernestus, Baayen, and Schreuder (2002) have shown that Dutch listeners recognize highly reduced word forms (taken from a corpus of spontaneous speech) only when such forms are presented in their full context (i.e., in a context of several words). When they are presented in isolation or in a very limited context, listeners do not recognize them. Apparently, not all possible phonetic variants of words are represented in the mental lexicon or, if they are, they are not equally accessible. Otherwise, recognition of reduced word forms in isolation would not be problematic. Furthermore, their findings suggest that hearing reduced word forms in context induces a process of reconstruction, which makes them difficult to distinguish from their non-reduced counterparts. The present study investigates whether reconstruction does indeed take place and — if so — what the precise nature of this reconstruction process might be.

Phonemic restoration is a powerful auditory illusion that was first reported by Warren (1970). Phonemic restoration typically occurs when part of an utterance is deleted and replaced with an extraneous sound (e.g., a cough or white noise). Listeners report that such an utterance sounds intact. They appear to 'hear' parts of words that are not really present in the acoustic signal (Samuel, 1996a). For phonemic restoration to be effective, the extraneous sound must have some spectral resemblance to the missing speech sound(s) (Samuel, 1981a, 1981b). Replacing speech sounds with silence does not lead to phonemic restoration (Warren, 1970).

Taft and Hambly (1985) showed that listeners might also restore reduced vowels to their full counterparts. In a syllable-monitoring task, listeners tended to accept a match between a target syllable and a word when the vowel of the target syllable was full while the vowel of the word was reduced, but only when the full vowel in question was consistent with the spelling of the reduced vowel. For example, while listeners detected [læg] in *lagoon* [ləguːn], they did not detect [lɔg] in the same word. [læg] and [lɔg] are both not present in the acoustic signal of *lagoon*, but [læg] is consistent with the spelling of *lagoon* whereas [lɔg] is not.

Casual reduced speech is fundamentally of a very different nature than the stimuli used in the traditional phonemic restoration studies, which tend to be carefully realized stimuli. In spontaneous reduced speech, phonemes are missing, but they are neither replaced by an extraneous sound nor by reduced phonemes. They are simply not realized. Their absence is inherent to the type of speech, and their deletion is systematic and highly frequent. Furthermore, the reductions in spontaneous speech are more dramatic than, for instance, the vowel reductions studied by Taft and Hambly (1985): Complete morphemes may be reduced to a single phoneme, or they may not be realized at all. Experiment 5.1 was designed to answer the question whether listeners might nevertheless restore missing speech sounds in highly reduced speech.

Experiment 5.1

We presented listeners with words ending in the highly frequent derivational suffix *-(e)lijk* [(ə)lək] ('-ly'). In spontaneous speech, this suffix is often severely reduced. For instance, *vreselijk* [fʀesələk] ('terrible') may be realized as [fʀesk], and *eigenlijk* [ɛiχələk] ('actually') may be realized as [ɛiχək]. In these examples, the phoneme [l] is present in the non-reduced realizations, but absent in the reduced realizations. We presented our participants with both non-reduced realizations (containing the phoneme [l]) and reduced realizations (not containing the phoneme [l]), and asked them to perform a phoneme-monitoring task, with the target phoneme [l]. We presented the target words in two conditions. In the first part of the experiment, we presented the target words in a context of several words (Full Context). In the second part of the experiment, we only presented the suffix or whatever was left of it (Minimal Context). In Full Context, participants were expected to recognize the reduced word forms (cf. Ernestus et al., 2002), and thus, restoration of missing phonemes on the basis of the activated non-reduced lexical representations of the target words should be possible. In this condition, our participants may report the presence of the phoneme [l], not only for the non-reduced forms, but also for the reduced forms. In other words, we expected many 'false positive' responses (i.e., reporting the presence of the phoneme [l] when this phoneme is not actually present in the acoustic signal). In Minimal Context, however, only small fragments of the target words were presented. As a result, the representations of the target words should not be activated and thus restoration of missing phonemes should not be possible. This condition enabled us to establish whether listeners are capable of accurately detecting the presence of the phoneme [l] when their perception cannot be influenced by the activated representations of the target words. Accurate detection performance in Minimal Context paired with many false positives in Full Context would constitute solid evidence for restoration of missing phonemes in reduced word forms.

Method

Participants. Fifty-one participants, mostly students at the University of Nijmegen, were paid to participate in the experiment. All were native speakers of Dutch.

Materials. We selected 11 Dutch words ending in the derivational suffix *-(e)lijk* [(ə)lək] ('-ly') as target stimuli. From a corpus of spontaneous speech (Ernestus, 2000), we selected both a non-reduced and a reduced realization of every target word (see Appendix A). In the non-reduced variants, the suffix was always fully realized. In the reduced variants, the suffix was either completely or partly reduced, but in no case was the phoneme [l] present in the realization (as established by two trained phoneticians).

The filler words in our experiment were 60 words, half of which contained a derivational affix other than *-(e)lijk* (e.g., *be-, ge-, ont-, -baar, -isch*), and half of which did not contain a derivational affix. Half of all filler words contained the phoneme [l]. For every filler word, two non-reduced tokens were selected from the corpus of spontaneous speech mentioned above. This resulted in 60 (non-reduced) realizations of filler words containing the phoneme [l], and 60 (non-reduced) realizations of filler words without the phoneme [l].

The target items and the filler items were put into a list and the order of presentation was pseudo-randomized three times (target items never occurred consecutively), resulting in three lists. Ten practice items were selected from the speech corpus: five non-reduced realizations of words containing the phoneme [l] and five non-reduced realizations of words without [l]. The order of presentation of the practice items was also randomized three times.

All items were presented in the two conditions. In Full Context (first part of the experiment), the items were presented in a context of several words. The context never contained the phoneme [l]. In Minimal Context (second part of the experiment), only the suffix *-(e)lijk* was presented for the non-reduced versions of the target words. For the reduced versions of the target words, we presented whatever was left of the suffix. For example, for the non-reduced realization of *eigenlijk* ([ɛiχələk˺]), we presented [ələk˺], and for the reduced realization of *eigenlijk* ([ɛiχək]), we presented [ək]. For the filler and the practice items containing the phoneme [l], we presented only the phoneme [l] plus approximately two to three surrounding phonemes. Finally, for the filler and the practice items without the phoneme [l], we presented a randomly selected portion of signal consisting of approximately two to three phonemes.

Procedure. Participants performed a phoneme-monitoring task, with the target phoneme [l]. They were instructed (on paper) to listen carefully to the stretches of speech that were presented to them, and to decide whether these stretches contained the sound [l]. The participants responded by pressing one of two buttons on a button box. Immediately after each button press, the participants indicated how confident they were about their decision, by circling one of the numbers on a rating scale ranging from 1 to 5 (1 = least confident, 5 = most confident). Each trial consisted of the presentation of a warning tone (377 Hz) for 500 ms, followed after an interval of 200 ms by the auditory stimulus. Stimuli were presented through Sennheiser headphones. After circling one of the numbers on the rating scale, the participants moved on to the next trial by pressing either button on the button box. The total experiment consisted of 284 trials: The first 142 trials constituted the Full Context Condition, the second 142 trials constituted the Minimal Context Condition. Prior to both conditions, the corresponding set of practice trials was presented, followed by a short pause. Within both conditions, there was a pause half-way through, that is, after 71 trials. The two conditions were separated by another pause. The total duration of the experimental session was approximately 30 minutes.

Results and discussion

In total, 2244 target trials were presented (11 target stimuli x 2 types x 2 conditions x 51 participants). Due to technical failure (the software not registering some button presses), the responses to only 2013 trials were recorded (i.e., the proportion of missing trials was 10.3%). Table 5.1 summarizes the percentages of 'yes'-responses ('Yes, I heard the sound [l]') to the non-reduced and the reduced target words, in both Full Context and Minimal Context. The percentages are calculated by dividing the number of 'yes'-responses in each cell of the design by the total number of responses in each cell.

Table 5.1: Percentages of 'yes'-responses to non-reduced and reduced word forms in Full Context and in Minimal Context in Experiment 5.1.

Type of word form	Full Context	Minimal Context
Non-reduced ([l] present)	87% (418 out of 482 trials)	84% (439 out of 524 trials)
Reduced (no [l] present)	*62% (308 out of 498 trials)*	*3% (16 out of 509 trials)*

The slanted percentages are the percentages of false positive responses (i.e., reporting the presence of the phoneme [l] when no [l] was actually present in the acoustic signal). The number of false positive responses was considerably higher in Full Context than in Minimal Context. A logistic regression analysis with the ratio of the numbers of 'yes'- and 'no'-responses for a given item as the dependent variable, and Type of word form (non-reduced versus reduced word form) and Context (Full Context versus Minimal Context) as factors yielded significant main effects of Type of word form ($\chi^2(1) = 620.7, p < 0.0001$) and Context ($\chi^2(1) = 299.8, p < 0.0001$), and a significant interaction of Type of word form and Context ($\chi^2(1) = 162.75, p < 0.0001$). Logistic regression analyses on the data for the two Context Conditions separately, revealed significant simple effects of Type of word form for both Context Conditions (Full Context: $\chi^2(1) = 81.8, p < 0.0001$; Minimal Context: $\chi^2(1) = 811.0, p < 0.0001$). The pattern in the interaction between Type of word form and Context shows that, when the critical stretches of speech (i.e., the non-reduced and the reduced suffixes) are presented outside their linguistic context, listeners are capable of accurately detecting the phoneme [l]. In contrast, when these stretches of speech are presented within their linguistic context, listeners tend to incorrectly report the presence of the phoneme [l] in the reduced word forms. In other words, listeners restore the missing phoneme [l].

For the analysis of the confidence scores (and for all following analyses of confidence scores), we pooled the scores for all 'yes'-responses over all items and all subjects. As the scores were not distributed normally, we analyzed them by means of Wilcoxon's Test. The average confidence scores of the 'yes'-responses in Full Context were not significantly different for the non-reduced and the reduced word forms (4.9 and 4.8 respectively, Wilcoxon's $W = 33379.5, p = 0.16$, two-tailed). This shows that the participants were equally confident about having heard the phoneme [l], whether or not this phoneme [l] was actually present in the acoustic signal. Apparently, the restoration escaped participants' awareness. This was supported by the participants' comments after the experiment.

Reaction times were measured from the onset of the first segment following the [l], or in the case of the reduced word forms, from the onset of the first segment that followed the position at which the [l] would have been realized. For the analysis of the reaction times (and for all following analyses of reaction times), we pooled the reaction time data for the 'yes'-responses over all items and all subjects, and we applied Wilcoxon's Test. This analysis showed that, although restoration of missing phonemes may occur at an unconscious level, it does take time. In Full Context, the reaction times corresponding to the false positives were significantly longer (242 ms on average) than the reaction times corresponding to the hits (i.e., correctly reporting the presence of [l]; Wilcoxon's $W = 84063, p < 0.0001$). The reaction times corresponding to the false positives were 1231 ms on average, and the reaction times corresponding to the hits were 989 ms on average. Our explanation for this finding is that the false positive responses are mediated at a lexical level of representation that becomes available at or after lexical access, whereas the hits reflect a phonetic level of processing (cf. Hallé, Chéreau, & Segui, 2000).

To summarize, listeners tend to restore phonemes that are missing in reduced word forms. Confidence scores suggest that listeners are not aware of the fact that they do. Nevertheless, the restoration takes time, probably because it is dependent on the lexical information that becomes available only after the intended word has been recognized.

Experiment 5.2

In the Minimal Context Condition in Experiment 5.1, we presented very short stretches of speech, only containing the suffix itself, such that the participants would not recognize the words. It is possible that a trace of a 'missing' [l] was present in the Full Context, for instance, in the form of a change in the quality of the preceding vowel, that was absent in the short stretches of speech presented in the Minimal Context. If so, then it is no wonder that participants more often responded 'no' in Minimal Context than in Full Context (where the whole form, including the possible [l]-trace, was presented to them). In order to establish whether traces of the [l] were indeed present in our reduced word forms, we ran a control experiment, employing the same phoneme-monitoring task as in Experiment 5.1. We presented the same target stimuli as in the Minimal Context Condition of Experiment 5.1, but now a larger portion of them was presented, so that any possible acoustic trace of [l] would be included.

Method

Participants. Ten students at the University of Nijmegen were paid to participate in the experiment. All were native speakers of Dutch. None of them had participated in Experiment 5.1.

Materials. The target stimuli in this experiment were the same 11 non-reduced and 11 reduced realizations as in Experiment 5.1. We now presented the suffix -*(e)lijk* plus the preceding vowel and any intervening consonant. For example, for the non-reduced realization of *vreselijk* [fʀeslək], we presented [eslək]. For the reduced realization of *vreselijk* [fʀesk], we presented [esk]. Any trace of the [l] left in the preceding vowel ([e]) is now included. Twenty-two non-reduced filler items were added, pseudo-randomly selected from the set of filler items of Experiment 5.1 (one realization per word). Half of these filler items contained the phoneme [l]. In addition, ten non-reduced practice items were presented, half of which contained an [l]. For the filler items and the practice items, we presented stretches of 4 to 5 phonemes.

The target items and the filler items were put into a list and the order of presentation was pseudo-randomized three times (the first three items were filler items, no more than two target items occurred consecutively, and the non-reduced and the reduced variants of one word never occurred consecutively). The order of presentation of the practice items was also randomized three times.

Procedure. The participants again performed a phoneme-monitoring task (target phoneme [l]). For the details of the procedure, see Experiment 5.1. The total experiment consisted of 10 practice trials and 44 experimental trials. The total duration of the experimental session was approximately 5 minutes.

Results and discussion

In total, 220 target trials were presented (11 target stimuli x 2 types x 10 participants). Due to technical failure, the response to one trial was not recorded (i.e., the proportion of missing trials was 0.5%).

For the non-reduced forms, 9 out of 11 forms received 100% ‘yes’-responses, indicating that the phoneme [l] was clearly perceivable in these forms. One form (*waarschijnlijk*) received 87% ‘yes’-responses and one form (*eerlijk*) received only 56% ‘yes’-responses. For the reduced forms, 8 out of 11 forms received 100%

'no'-responses, indicating that these forms did not contain the phoneme [l] or any perceivable trace of it. One form (*mogelijk*) received 56% 'no'-responses, one form (*duidelijk*) received 33% 'no'-responses, and one form (*onmiddellijk*) received 11% 'no'-responses. Clearly, these three forms contained some acoustic trace leading the participants to perceive the phoneme [l].

Given these results, we re-analyzed the data of Experiment 5.1, now excluding the non-reduced form of *eerlijk* and the reduced forms of *mogelijk, duidelijk* and *onmiddellijk* from both the Full Context Condition and the Minimal Context Condition. Table 5.2 summarizes the resulting percentages of 'yes'-responses to the non-reduced and the reduced target words, in both Full Context and Minimal Context. The percentages are calculated by dividing the number of 'yes'-responses in each cell of the design by the total number of responses in each cell.

Table 5.2: Percentages of 'yes'-responses to non-reduced and reduced word forms in Full Context and in Minimal Context in Experiment 5.1, after exclusion of the non-reduced form of *eerlijk* and the reduced forms of *mogelijk, duidelijk*, and *onmiddellijk*.

Type of word form	Full Context	Minimal Context
Non-reduced ([l] present)	87% (378 out of 434 trials)	91% (435 out of 479 trials)
Reduced (no [l] present)	*52% (184 out of 352 trials)*	*2% (7 out of 373 trials)*

The number of false positive responses was still considerably higher in Full Context than in Minimal Context (slanted percentages). A logistic regression analysis with the ratio of the numbers of 'yes'- and 'no'-responses for a given item as the dependent variable, and Type of word form (non-reduced versus reduced word form) and Context (Full Context versus Minimal Context) as factors again yielded significant main effects of Type of word form ($\chi^2(2) = 721.3, p < 0.0001$) and Context ($\chi^2(1) = 126.3, p < 0.0001$), and a significant interaction of Type of word form and Context ($\chi^2(1) = 156.3, p < 0.0001$). Logistic regression analyses on the data for the two Context Conditions separately, yielded significant simple effects of Type of word form for both conditions (Full Context: $\chi^2(1) = 118.4, p < 0.0001$; Minimal Context: $\chi^2(1) = 816.5, p < 0.0001$). In other words, the analyses of the response data without the problematic items led to similar results as the original analyses.

The same holds for the analysis of the confidence scores. In Full Context, the average confidence scores for the 'yes'-responses did not differ for the non-reduced word forms and the reduced word forms (4.9 and 4.8 respectively; Wilcoxon's $W = 33379.5, p = 0.16$, two-tailed). In addition, the reaction times were still longer (448 ms on average) for the false positives than for the hits (Wilcoxon's $W = 54190, p <$

0.0001, two-tailed). The reaction times corresponding to the false positives were 1420 ms on average, and the reaction times corresponding to the hits were 971 ms on average.

In summary, listeners are able to detect accurately the presence of the phoneme [l] in phoneme strings that are presented without any linguistic context. When the same phoneme strings are presented within their natural context, listeners frequently incorrectly report the presence of the phoneme [l] in reduced word forms, just as confidently as when they correctly report the presence of the phoneme [l] in non-reduced word forms. In other words, listeners restore the missing phoneme [l].

Experiment 5.3

Restoration is based on the information that becomes available once a lexical representation is accessed (Warren, 1970; Samuel, 1981a, 1987, 1996b). This information may be phonological or orthographic in nature. Previous studies have shown that orthographic information is influential in phoneme-monitoring experiments (Seidenberg & Tanenhaus, 1979; Donnenwerth-Nolan, Tanenhaus, & Seidenberg, 1981; Taft & Hambly, 1985; Dijkstra, Roelofs, & Fieuws, 1995; Treiman & Cassar, 1997; Hallé et al., 2000), especially when the experiment contains many words with deviant spellings (Cutler, Treiman, & Van Ooijen, 1998). All these experiments concerned careful laboratory speech, which is relatively slow compared to natural, spontaneous speech. It is not clear what the nature of the lexically provided information may be that results in the restoration of highly reduced word forms in spontaneous speech, as observed in Experiment 5.1. Is the restoration the result of the activation of the phonological code of the word, is it the result of the activation of the orthographic code, or is it the result of the activation of both?

An answer to this question may be found in a comparison of the results of our first two experiments with the results of a similar experiment in which restoration, if at all, must necessarily take place on the basis of orthographic information. In Experiment 5.3, we presented our participants with items that either did or did not contain a mismatch between the orthographic and the phonological code. The matching items were words for which the orthographic code contained the digraph *ei/ij* (the two possible spellings of [ɛi] in Dutch) and the phonological code contained the phoneme [ɛi] (e.g., *bijna* [bɛinɑ] ‘almost’). The mismatching items were words ending in the suffix *-(e)lijk*. This suffix forms one of the few exceptions to the generalization that the digraph *ij* in Dutch is realized as [ɛi], as, in this suffix, the digraph *ij* is always pronounced as [ə]. Thus, in words ending in *-(e)lijk*, there is a mismatch between the orthographic code and the phonological code. Since [ɛi] is pronounced as [ə] in only very few morphemes, this mismatch is more severe than the vowel reductions studied by Taft and Hambly (1985).

Again, the participants performed a phoneme-monitoring task, but now the target phoneme was [ɛi] instead of [l]. As in Experiment 5.1, all items were presented in two conditions. In the Full Context Condition, the items were presented in a context of several words. If the restoration phenomenon that we have observed in the corresponding condition in Experiment 5.1 were (at least in part) the result of the activation of orthographic codes, we would expect our participants to (incorrectly) report the presence of the phoneme [ɛi] in the words with the suffix *-(e)lijk* in this condition. In contrast, if the restoration phenomenon were solely the result of the activation of phonological codes, we would expect our participants to (correctly) report the absence of the phoneme [ɛi] in the words with the suffix *-(e)lijk*.

In the Minimal Context Condition, we presented only the phoneme corresponding to the digraph *ij/ei* (i.e., [ə] for the words with *-(e)lijk* and [ɛi] for all other words) plus two to three surrounding phonemes. In this condition, we expect participants to accurately detect the presence of the phoneme [ɛi]. If restoration of phonemes occurs on the basis of activated orthographic representations, we may expect accurate detection performance in Minimal Context paired with many false positive responses in Full Context.

Method

Participants. Forty-seven participants, mostly students at the University of Nijmegen, were paid to participate in the experiment. All were native speakers of Dutch. None of them had participated in Experiment 5.1 or 5.2.

Materials. From the corpus of spontaneous speech (Ernestus, 2000), we selected 22 non-reduced realizations of words with the suffix *-(e)lijk* [(ə)lək], that is, words with a mismatch between the orthographic code and the phonological code (see Appendix B). Additionally, we selected 22 non-reduced realizations of words in which there is no such mismatch: words of which the orthographic code contains the digraph *ei* or *ij* and the phonological code contains the diphone [ɛi] (e.g., *bijna* [bɛina], see Appendix B).

Sixty-four filler items were added: 20 non-reduced realizations of words containing both *ij/ei* and [ɛi], with derivational affixes other than *-(e)lijk* (e.g., *be-, ge-, ont-,*

-baar, -isch); 22 non-reduced realizations of words not containing *ij/ei* or [ɛi], with derivational affixes other than *-(e)lijk*; and 22 non-reduced realizations of words, without derivational affixes, not containing *ij/ei* or [ɛi]. Twenty practice items were selected, half of which contained both *ei/ij* and [ɛi] and half of which did not contain *ei/ij* or [ɛi]. The target items and the filler items were put into a list and the order of presentation was pseudo-randomized three times (no more than two target items occurred consecutively), resulting in three lists. The order of presentation of the practice items was also randomized three times.

In the Full Context Condition (first part of the experiment), the items were presented in a context of several words. In the Minimal Context Condition (second part of the experiment), we presented only the phoneme corresponding to the digraph *ij/ei* (i.e., [ə] for the words with *-(e)lijk* and [ɛi] for all other words) plus two to three surrounding phonemes. For example, for the realization *eigenlijk* [ɛiχələk˺], we presented [ələk˺], and for the realization *bijna* [bɛina] ('almost'), we presented [bɛin]. For the items without the digraph *ij/ei*, a randomly selected portion of signal consisting of two to three phonemes was presented.

Procedure. Participants performed a phoneme-monitoring task, with the target phoneme [ɛi]. For the details of the procedure, see Experiment 5.1. In contrast to Experiment 5.1, the participants were instructed orally. Moreover, as in Experiment 5.1, the confidence rating scale ranged from 1 to 5 (1 = least confident, 5 = most confident), but now participants were also given the option to indicate that they had given the wrong response, by circling "F". The total experiment consisted of 216 trials: The first 108 trials constituted the Full Context Condition, the second 108 trials constituted the Minimal Context Condition. Both conditions were presented in three blocks of 36 trials, separated by short pauses. Prior to each condition, the corresponding set of practice trials was presented, followed by a short pause. The total duration of the experimental session was approximately 30 minutes.

Results and discussion

In total, 4136 target trials were presented (22 target stimuli x 2 types x 2 conditions x 47 participants). We excluded those trials from the analyses for which participants circled "F" on the rating scale form (143 trials, 3.5%). Due to technical failure, another 599 responses (14.5%) were missing. Table 5.3 summarizes the remaining percentages of 'yes'-responses to the matching word forms (both *ij/ei* and [ɛi] present) and to the mismatching word forms (*ij* but not [ɛi] present), in both Full

Context and Minimal Context. The percentages are calculated by dividing the number of 'yes'-responses in each cell of the design by the total number of responses in each cell.

Table 5.3: Percentages of 'yes'-responses to matching word forms (containing *ij/ei* and [ɛi]) and mismatching word forms (containing *ij* but not [ɛi]) in Full Context and in Minimal Context in Experiment 5.3.

Type of word form	Full Context	Minimal Context
Matching ([ɛi] present)	91% (745 out of 821 trials)	90% (800 out of 884 trials)
Mismatching (no [ɛi] present)	*47% (400 out of 843 trials)*	*28% (241 out of 846 trials)*

The slanted percentages are the percentages of false positives (i.e., reporting the presence of the phoneme [ɛi] when no [ɛi] was actually present in the acoustic signal). Although we observed surprisingly many false positive responses in Minimal Context, the number of false positive responses was considerably higher in Full Context. A logistic regression analysis with the ratio of the numbers of 'yes'- and 'no'-responses for a given item as the dependent variable, and Type of word form (matching versus mismatching word form) and Context (Full Context versus Minimal Context) as factors yielded significant main effects of Type of word form ($\chi^2(2) = 1120.6, p < 0.0001$) and Context ($\chi^2(1) = 48.0, p < 0.0001$), and, importantly, a significant interaction of Type of word form and Context ($\chi^2(1) = 15.8, p < 0.0001$). Logistic regression analyses on the data for the two Context Conditions separately, revealed significant simple effects of Type of word form for both Context Conditions (Full Context: $\chi^2(1) = 396.0, p < 0.0001$; Minimal Context: $\chi^2(1) = 760.0, p < 0.0001$). The pattern in the interaction between Type of word form and Context shows that, when the critical stretches of speech are presented outside their linguistic context, listeners are more or less capable of detecting the phoneme [ɛi]. In contrast, they produce many false positives when the stretches of speech are presented in context. This shows that listeners base their responses at least partly on the orthographic representations of the words.

In fact, after the experiment, many participants reported that they had been aware of the presence of items that contained a mismatch between the phonological and the orthographic code, and that they had chosen to base their decisions on orthography. They had adopted the conscious decision strategy of reporting the phoneme [ɛi] whenever the orthographic code contained the digraph *ij* or *ei*.

This decision strategy is also reflected in the reaction times, which were not different for the false positive responses and the correct 'yes'-responses in this experiment (Wilcoxon's $W = 157094.5, p = 0.13$, two-tailed). The reaction times corresponding to the false positives were 1197 ms on average, and the reaction times corresponding to the hits were 1095 ms on average. This contrasts with Experiment 5.1, in which the false positive responses were significantly slower than the correct ‘yes’-responses. Apparently, the listeners in Experiment 5.3 always based their decision on lexical orthographic information, irrespective of stimulus type.

The average confidence scores for the ‘yes’-responses in Full Context were significantly lower for the mismatching items (i.e., for the items with *-(e)lijk*) than for the matching items (4.7 and 5.0 respectively; Wilcoxon’s $W = 114727.5, p < 0.0001$, two-tailed). This reflects the participants’ awareness of the mismatches between orthography and the presented acoustic realizations. In other words, although many participants used an orthography-based strategy, the actual realizations still affected the confidence scores.

The considerable number of false positive responses in Minimal Context suggests that, also in this condition, our listeners recognized the phoneme string [(ə)lək] as the suffix *-(e)lijk*. In other words, also in Minimal Context, the orthographic code (containing *ij*) appears to have occasionally been activated, and to have led the participants to report the presence of the phoneme [ɛi].

The participants probably applied the orthography-based strategy because the instructions had not mentioned possible mismatches between orthography and acoustic realizations, and the participants were therefore uncertain how to deal with such mismatches. The reason that they consciously relied more on orthography than on the acoustic signal may be that spontaneous speech is relatively fast and poorly intelligible, introducing uncertainty about whether certain segments are present.

Experiment 5.4

In Experiment 5.1, as opposed to in Experiment 5.3, there was no indication of the use of a strategy, neither in reaction times nor in participants’ comments. The results of Experiment 5.3 are therefore not informative about whether the restoration in Experiment 5.1 may have occurred on the basis of orthography. We carried out a fourth experiment, investigating whether orthography has an influence even when the instructions are explicit about the orthography-phonology mismatches and how to deal with such mismatches, that is, when it is emphasized that participants’ decisions should be based on the acoustic signal. In this experiment, participants were instructed to listen carefully to the acoustic signal, and if they did not hear [ɛi], as would be the case for all mismatching forms, they were to respond 'no', even if the orthography contained the digraph *ij*. In order to prevent participants from now adopting the strategy of responding 'no' whenever they heard [(ə)lək], without carefully listening to the acoustic signal, we included as fillers a number of catch trials: trials including the suffix *-(e)lijk* (suffix without [ɛi]), but with the phoneme [ɛi] at another position in the word (e.g., *eigenlijk* [ɛiχələk]) or in the context, so that participants would occasionally have to respond 'yes' to trials containing [(ə)lək]. Finally, it was stressed that the participants had to respond as quickly as possible. If orthography is difficult to ignore, we may expect relatively many false positive responses in the Full Context Condition of this experiment, despite the emphasis on listening to the acoustic signal and responding as fast as possible. Restoration of reduced word forms must then indeed be a consequence, at least in part, of the automatic activation of orthographic information.

Method

Participants. Twenty-one participants, mostly students at the University of Nijmegen, were paid to participate in the experiment. All were native speakers of Dutch. None of them had participated in Experiment 5.1, 5.2, or 5.3.

Materials. Exactly the same materials were used as in Experiment 5.3, plus 53 additional items (all taken from the corpus of spontaneous speech). Of these 53 additional items, 28 functioned as catch trials (in the Full Context Condition) in order to prevent participants from adopting the decision strategy "if *-(e)lijk* [(ə)lək], then 'no'-button". These items contained the suffix *-(e)lijk* and the phoneme [ɛi] at some other position in the word or in the context. The other 25 additional items consisted of 21 tokens of *bijvoorbeeld* ('for example'), in which *ij* was realized as [ə], and 4 tokens of *bijzonder* ('special'), with [i] as the first vowel. These items reminded the participants to press the 'yes'-button only if they heard [ɛi]. Two tokens of *bijvoorbeeld* and one token of *bijzonder* were included in the practice set. The additional items were evenly distributed over the trial lists that were used in Experiment 5.3, with the restriction that if the same word ending in *-(e)lijk* occurred both in a target trial and in a catch trial, the target trial was always presented first (e.g., the word *natuurlijk* occurred first in a context without [ɛi], and only later in the list in a context with [ɛi]).

As in Experiment 5.1 and 5.3, all items were presented in two conditions. In the first part of the experiment (Full Context Condition), the items were presented in a context of several words. In the second part of the experiment (Minimal Context Condition), we presented only the phoneme corresponding to the digraph *ij/ei* plus two or three surrounding phonemes. If neither the word nor the context contained *ij/ei*, we presented a randomly chosen phoneme sequence.

Procedure. Participants performed exactly the same task as in Experiment 5.3: a phoneme-monitoring task (target phoneme [ɛi]), with every trial followed by a confidence rating. For the details of the procedure, see Experiment 5.3. Again, the participants were instructed orally, but now special emphasis was laid on listening carefully to the acoustic signal and absolutely ignoring the orthography, and on responding as quickly as possible. Participants were provided with examples of words containing the digraph *ij/ei* without containing the phoneme [ɛi], such as *bijzonder* ([bizɔndəʀ]) and *bijvoorbeeld* ([bəvoːʀbeːlt]). Moreover, one of the examples was a word containing the suffix *-(e)lijk*. Participants were specifically instructed to respond ‘no’ to such words.

The total experiment consisted of 316 trials: The first 158 trials constituted the Full Context Condition, the second 158 trials constituted the Minimal Context Condition. Both conditions were presented in three blocks of approximately equal length, separated by short pauses. Prior to both the Full Context Condition and the Minimal Context Condition, the set of corresponding practice trials was presented, followed by a short pause. The total duration of the experimental session was approximately 45 minutes.

Results and discussion

In total, 1848 target trials were presented (22 target stimuli x 2 types x 2 conditions x 21 participants). Due to technical failure, for one participant the data of the Minimal Context Condition were missing, and for another participant the data of the Full Context Condition were missing. Distributed over the other participants, another 26 responses were missing. This amounted to 114 missing trials (6.2%). Furthermore, we excluded those trials from the analyses for which participants circled “F” on the rating scale form (100 trials, 5.4%). Table 5.4 summarizes the remaining percentages of ‘yes’-responses to the matching word forms (both *ij/ei* and [ɛi] present) and to the mismatching word forms (*ij* but not [ɛi] present), in both Full Context and Minimal Context. The percentages are calculated by dividing the number of ‘yes’-responses in each cell of the design by the total number of responses in each cell.

Table 5.4: Percentages of 'yes'-responses to matching word forms (containing *ij/ei* and [ɛi]) and mismatching word forms (containing *ij* but not [ɛi]) in Full Context and in Minimal Context in Experiment 5.4.

Type of word form	Full Context	Minimal Context
Matching ([ɛi] present)	92% (366 out of 396 trials)	89% (389 out of 436)
Mismatching (no [ɛi] present)	*13% (53 out of 394 trials)*	*5% (19 out of 408 trials)*

The slanted percentages are the numbers of false positive responses. Although the overall proportion of false positive responses was considerably lower than in Experiments 5.1 and 5.3, the number of false positive responses in Full Context was again higher than in Minimal Context [1]. A logistic regression analysis with the ratio of the numbers of 'yes'- and 'no'-responses for a given item as the dependent variable, and Type of word form (matching versus mismatching word form) and Context (Full Context versus Minimal Context) as factors yielded a significant main effect of Type of word form ($\chi^2(2) = 1268.8, p < 0.0001$), a significant main effect of Context ($\chi^2(1) = 17.8, p < 0.0001$), and a significant interaction of Type of word form and Context ($\chi^2(1) = 4.4, p < 0.05$). Logistic regression analyses on the data for the two Context Conditions separately, revealed significant simple effects of Type of word form for both Context Conditions (Full Context: $\chi^2(1) = 569.6, p < 0.0001$; Minimal Context: $\chi^2(1) = 717.3, p < 0.0001$). Within the dataset for the mismatching word forms, there was a significant effect of Context ($\chi^2(1) = 19.6, p < 0.0001$). These results suggests that, in Full Context, orthographic codes are activated automatically, and that, despite specific instructions to ignore the spelling of the words, participants' responses are still partly based on these orthographic codes.

However, a closer look at the response data in Experiment 5.4 shows that the processes underlying the overt response behaviour may not have been the same for all participants in this experiment. Only three participants were responsible for 65% of all false positive responses. The remaining 35% of the false positive responses were more or less equally distributed over the other participants. We ran separate analyses of the confidence scores and the reaction times for these two groups of participants. These analyses showed that, for the three participants with relatively many false positive responses, the pattern in the reaction times was similar to that in Experiment 5.3 (no difference in reaction times between hits (1334 ms) and false positives (1424 ms) in Full Context; Wilcoxon's $W = 708, p = 0.25$,

[1] Since the numbers of false positive responses were small, it is important to point out here that, in both conditions, the false positive responses were distributed evenly over a large number of different items, that is, they were not restricted to a small number of peculiar items.

two-tailed), while the pattern in the confidence ratings was similar to that in Experiment 5.1 (no difference in confidence ratings between the false positives and the hits in Full Context; Wilcoxon's $W = 865, p = 0.75$, two-tailed). We think that these three exceptional participants applied the same strategy as the participants in Experiment 5.3, that is, scanning the orthographic code for *ij/ei*, although they were explicitly instructed not to do so. They were more confident in their false positive responses than the participants in Experiment 5.3, basically because they ignored the instructions altogether.

For the other participants, we obtained the opposite result. Their pattern in the reaction times was similar to that in Experiment 5.1 (slower reaction times — 537 ms on average — for false positives (1523 ms) than for hits (986 ms) in Full Context; Wilcoxon's $W = 1253, p < 0.01$, two-tailed), while their pattern in the confidence ratings was similar to that in Experiment 5.3 (lower confidence scores for the false positives than for the hits — 4.1 and 5.0 respectively — in Full Context; Wilcoxon's $W = 1473.5, p < 0.0001$, two-tailed). This pattern in the reaction time data seems to suggest that the false positive responses by these participants were not the result of a conscious decision strategy, but, instead, that they were the result of unconscious restoration. If so, this restoration must necessarily have taken place on the basis of orthographic information. While the on-line responses of these participants are suggestive of the occurrence of restoration, the off-line confidence ratings suggest that, in retrospect, participants have been aware of the mismatching nature of the items that triggered the false positives (contrary to in Experiment 5.1). This is easily explained by the fact that the instructions for Experiment 5.4 were explicitly aimed at making the participants aware of the orthography-phonology mismatch.

Analyses of the data after exclusion of the three exceptional participants, however, show that, even though numerically the number of false positive responses was somewhat higher in Full Context than in Minimal Context (5% versus 2%), this difference was not statistically significant. A logistic regression analysis with the ratio of the numbers of 'yes'- and 'no'-responses for a given item as the dependent variable, and Type of word form (matching versus mismatching word form) and Context (Full Context versus Minimal Context) as factors, yielded a significant effect of Type of word form ($\chi^2(1) = 1326.7, p < 0.0001$), a marginally significant effect of Context ($\chi^2(1) = 3.7, p = 0.06$), and no interaction of Type of word form by Context ($\chi^2(1) = 0.4, p = 0.54$). Within the dataset for the mismatching word forms, there was no significant effect of Context ($\chi^2(1) = 2.4, p = 0.12$).

To conclude, in the case of a mismatch between orthography and phonology, receiving only very general instructions (Experiment 5.3) leads to great uncertainty (as shown by the confidence scores and the participants' comments) and, as a result, to the use of a conscious decision strategy (as suggested by the reaction times and reported by the participants). Receiving adequate, specific instructions (Experiment 5.4) reduces uncertainty about the task, but the mismatching nature of the items eliciting the false positives still affects the (off-line) confidence scores. Finally, the results of Experiment 5.4 suggest that the phonemic restoration of reduced word forms, as it took place in Experiment 5.1, occurs mainly on the basis of phonological information — even though there may be a small role for orthography (as suggested by the reaction time data and the slightly (but statistically non-significantly) higher number of false positive responses in Full Context than in Minimal Context in the dataset without the three exceptional participants).

General discussion

In the first two experiments of this study, we investigated the processing of highly reduced word forms, as they occur in casual spoken Dutch. In these reduced forms, suffixes may be either completely or partly missing. Highly reduced word forms are not recognized when they are presented in isolation, whereas they are recognized when they occur in their natural context (Ernestus et al., 2002). A possible explanation for this finding may be that when highly reduced forms are recognized in their natural context, restoration of partly or completely missing suffixes takes place. The results of our first two experiments confirm this hypothesis. Listeners report the presence of a missing phoneme [l] in reduced forms, but only when these forms are presented in a context of several words, that is, only when the reduced forms can be recognized. When the critical stretches of phonemes are presented in isolation, listeners accurately discriminate items containing [l] and items not containing [l].

In the classic phonemic restoration studies, restoration has been shown to occur for carefully realized stimuli in which one or more phonemes have been replaced by a (spectrally similar) sound. Our experiments show that partly or completely missing suffixes are restored in naturally reduced speech, when the missing phonemes are not replaced by another sound and a large portion of the word is missing. Highly reduced forms like the ones studied in these experiments occur frequently in casual spoken Dutch. This suggests that restoration is a natural, highly frequent process.

Experiments 5.3 and 5.4 investigated the nature of the lexically provided information that the restoration of reduced word forms is based on. Does restoration occur on the basis of phonology, orthography, or both? The results of Experiments 5.3 and 5.4 suggest that orthographic information may play a role: When presented with stimuli with mismatching orthography and phonology, and when explicitly asked to ignore orthography, listeners still occasionally base their responses on the orthographic information. This suggests that orthography is activated automatically when a word is recognized in a phoneme-monitoring task, and that it is difficult to ignore. This is in line with other evidence that phonological processing may be influenced by orthographic representations (Seidenberg & Tanenhaus, 1979; Donnenwerth-Nolan et al., 1981; Taft & Hambly, 1985; Dijkstra et al., 1995; Treiman & Cassar, 1997; Hallé et al., 2000). The relatively small number of false positive responses in Experiment 5.4 as compared to Experiment 5.1, however, suggests that, although orthography may play a role, phonology is the main source of information that the restoration in Experiment 5.1 was based on.

To conclude, previous research has shown that listeners cannot recognize highly reduced word forms when they are presented in isolation (Ernestus et al., 2002). The present study sheds more light on this issue, by investigating the processing of casually realized word forms, in which the suffix *-(e)lijk* is either partly or completely reduced. By means of a traditional phoneme-monitoring task, we showed that, when these reduced word forms are presented in context, the suffixes that are missing in these forms are restored: Listeners 'hear' the suffixes that are missing. The conscious percept is based on the activated canonical representation, not so much on (a pre-lexical representation of) the acoustic signal itself: The activated representations in the lexicon determine what we think we hear. In isolation, restoration does not occur. As a consequence, listeners may not recognize reduced word forms when presented in isolation.

The restoration phenomenon shows that the reduced forms are linked to the canonical representations in the mental lexicon, and that the canonical, non-reduced representations (as well as the corresponding orthographic representations) are highly activated upon hearing reduced forms. This finding has implications for episodic models assuming that all surface forms of words are stored in the mental lexicon (e.g., Goldinger, 1998). Even though, in such an architecture, lexical access is mediated by a representation of a reduced form, the non-reduced word form reaches the highest level of activation. It overrules the activation of the representation of the actual reduced form, which results in restoration: The phoneme-monitoring response cannot be executed on the basis of the reduced representation. Activation of the canonical representation is consistent with models that assume that all words are represented by just one representation in the lexicon. These models, however, face the challenge of how to map a highly reduced word form such as [tyk] onto the non-reduced canonical representation ([natyʀlək] - *natuurlijk*).

Restoration enables listeners to understand spontaneous speech, without comprehension being hampered at a conscious level by the drastic reductions inherent to this type of speech. The restoration of highly reduced word forms may partly be based on orthographic information, but phonological information appears to be most important.

References

Cutler, A., Treiman, R., and Van Ooijen, B.: 1998, Orthografik inkonsistensy ephekts in foneme detektion?, *Proceedings of the Fifth International Conference on Spoken Language Processing* **6**, 2783–2786.

Dijkstra, A., Roelofs, A., and Fieuws, S.: 1995, Orthographic effects on phoneme monitoring, *Canadian Journal of Experimental Psychology* **49**, 264–271.

Donnenwerth-Nolan, S., Tanenhaus, M., and Seidenberg, M.: 1981, Multiple code activation in word recognition: Evidence from rhyme monitoring, *Journal of Experimental Psychology: Human Learning and Memory* **7**, 170–180.

Ernestus, M.: 2000, *Voice assimilation and segment reduction in casual Dutch, a corpus-based study of the phonology-phonetics interface*, LOT, Utrecht.

Ernestus, M., Baayen, R. H., and Schreuder, R.: 2002, The recognition of reduced word forms, *Brain and Language* **81**, 162–173.

Goldinger, S. D.: 1998, Echoes of echoes? An episodic theory of lexical access, *Psychological Review* **105**, 251–279.

Hallé, P., Chéreau, C., and Segui, J.: 2000, Where is the /b/ in ”absurde” [apsyʀd]? It is in French listeners’ minds, *Journal of Memory and Language* **43**, 618–639.

Samuel, A.: 1981a, Phonemic restoration: Insights from a new methodology, *Journal of Experimental Psychology: General* **110**, 474–494.

Samuel, A.: 1981b, The role of bottom-up confirmation in the phonemic restoration illusion, *Journal of Experimental Psychology: Human Perception and Performance* **7**, 1124–1131.

Samuel, A.: 1987, Lexical uniqueness effects on phonemic restoration, *Journal of Memory and Language* **26**, 36–56.

Samuel, A.: 1996a, Does lexical information influence the perceptual restoration of phonemes?, *Journal of Experimental Psychology: General* **125**, 28–51.

Samuel, A.: 1996b, Phoneme restoration, *Language and Cognitive Processes* **11**, 647–653.

Seidenberg, M. S. and Tanenhaus, M. K.: 1979, Orthographic effects on rhyme monitoring, *Journal of Experimental Psychology: Human Learning and Memory* **5**, 546–554.

Taft, M. and Hambly, G.: 1985, The influence of orthography on phonological representations in the lexicon, *Journal of Memory and Language* **24**, 320–335.

Treiman, R. and Cassar, M.: 1997, Can children and adults focus on sounds as opposed to spelling in a phoneme counting task?, *Developmental Psychology* **33**, 771–780.

Warren, R.: 1970, Perceptual restoration of missing speech sounds, *Science* **167**, 392–393.

Appendix A

Non-reduced and reduced variants of the target words in Experiment 5.1:

Word	Non-reduced form	Reduced form
duidelijk ‘clear(ly)’	[dœydlək]	[dœyɾə]
eerlijk ‘honest(ly)’	[ɪːlək˺]	[ɪː]
eigenlijk ‘actual(ly)’	[ɛiχələk˺]	[ɛiχək]
mogelijk ‘possible/possibly’	[moχələk]	[moχəg˺]
namelijk ‘namely’	[namələk]	[namg]
natuurlijk ‘of course’	[nətylək]	[tyk]
onmiddellijk ‘immediate(ly)’	[ɔ̃mɪdələk]	[ɔ̃mɪdək˺]
uiteindelijk ‘eventual(ly)’	[œytɛ̃ilələg˺]	[œytɛiŋ]
verschrikkelijk ‘terrible/terribly’	[fəsʀɪklək]	[fsɪk]
vreselijk ‘terrible/terribly’	[fʀeslək]	[fʀesk]
waarschijnlijk ‘probable/probably’	[ʋᵊsχɛ̃lək˺]	[fsχɛŋk˺]

Appendix B

Target items in Experiment 5.3:

22 Items with the suffix *-(e)lijk* [(ə)lək] (*ij* in the orthographic code, but no [ɛi] in the phonological code): *afhankelijk* 'dependent', *afzonderlijk* 'separate', *behoorlijk* 'decent', *duidelijk* 'clear', *eerlijk* 'honest', *heerlijk* 'delightful', *makkelijk* 'easy', *moeilijk* 'difficult', *mogelijk* 'possible', *nadrukkelijk* 'emphatic', *namelijk* 'namely', *natuurlijk* 'of course', *onmiddellijk* 'immediate', *onredelijk* 'unreasonable', *ontzaglijk* 'immense', *oostelijk* 'eastern', *persoonlijk* 'personal', *redelijk* 'reasonable', *verantwoordelijk* 'responsible', *verschrikkelijk* 'awful', *vreselijk* 'awful', *vrolijk* 'cheerful'.

22 Items with *ei/ij* in the orthographic code and [ɛi] in the phonological code: *beide* 'both', *bijna* 'almost', *blij* 'happy', *blijkt* 'appears', *blijven* 'to stay', *eigen* 'own', *einde* 'end', *kijken* 'to look', *klein* 'small', *kwijt* 'lost', *lijkt* 'seems', *partij* 'party', *pijn* 'pain', *prijs* 'price', *slijten* 'to wear out', *termijn* 'term', *trein* 'train', *vrij* 'free', *weinig* 'few', *wijk* 'neighbourhood', *zei* 'said', *zijn* 'to be'.

Summary and conclusions

CHAPTER 6

This thesis investigated two seemingly contradictory properties of the speech-perception system. On the one hand, listeners are extremely sensitive to the fine phonetic details in the speech signal. These subtle acoustic cues can reduce the temporal ambiguity between words that show initial segmental overlap, and can guide lexical activation (e.g., Davis, Marslen-Wilson, & Gaskell, 2002; Spinelli, McQueen, & Cutler, 2003; Salverda, Dahan, & McQueen, 2003). On the other hand, comprehension does not seem to be hampered at all by the drastic reductions that typically occur in casual speech. Complete segments, and sometimes even complete syllables, may be missing, but comprehension is seemingly unaffected (Ernestus, 2000; Ernestus, Baayen, & Schreuder, 2002). This thesis aimed at elucidating how words are represented and accessed in the mental lexicon, by investigating these contradictory phenomena for the domain of morphology. Chapters 2 through 4 investigated the role of subsegmental cues in the processing of words that show segmental overlap *and* that are morphologically — inflectionally or derivationally — related. Chapter 5 studied the processing of highly reduced forms, in particular Dutch words in which the derivational suffix *-(e)lijk* [(ə)lək] ('-ly') has been drastically reduced.

Subsegmental differences between stems in isolation and stems in morphologically complex words

A first question addressed in this thesis is whether the subsegmental differences observed in earlier studies between the initial syllables of short words and segmentally overlapping, but morphologically *unrelated* longer words (e.g., *cap* - *captain*) are also consistently present between the initial syllables of short words and segmentally overlapping, morphologically *related* longer words (e.g., *work - worker*). The morphological relations studied in this thesis are regular inflection and derivation, in Dutch and in English. In particular, we compared stems occurring in isolation with stems that are followed by an inflectional or derivational suffix. As inflectional and derivational suffixes in both Dutch and English often consist of only one or two segments (e.g., the Dutch plural suffix *-en* is usually realized as just a schwa), it is not self-evident that the addition of a suffix would lead to substantial acoustic changes in the stem. Chapters 2, 3, and 4 showed that it does.

Chapter 2 investigated whether there are durational and intonational differences between singular nouns and the stems of plural nouns in Dutch. The nouns under investigation were monosyllabic nouns that take the plural suffix *-en* [ə(n)] (e.g., *boek - boeken*, [buk] - [bukə(n)], 'book' - 'books'). One speaker's realizations of the singular and plural forms of these nouns were studied. Acoustic measurements on the noun forms realized in isolation showed that the stem of the plural form (which is segmentally identical to the singular form) was on average acoustically shorter than the singular form, and that it had a higher average fundamental frequency than the singular form.

In Chapter 3, similar durational differences were observed between monosyllabic adjectives and the stems of their corresponding comparatives (e.g., *strict - stricter*), and between monosyllabic verb stems and the stems of their corresponding agent nouns (e.g., *work - worker*). These durational differences were observed for Dutch as well as for English (one speaker per language).

Chapter 4 investigated whether similar durational differences are present between singular forms and the stems of plural forms for Dutch nouns that take the plural suffix *-s* [s], even though the addition of the [s] does not entail the addition of an extra syllable. The singular and plural nouns were embedded in non-ambiguous sentences (e.g., *Ik spreek de kerel* (singular)/*kerels* (plural) *vaak*, 'I often talk to the guy/guys') and in ambiguous sentences (e.g., *Ik spreek de kerel/kerels soms*, 'I sometimes talk to the guy/guys'). The ambiguous sentences always contained a noun followed by a word with an [s]-onset. As a result of degemination of the cluster of *s*-es, a plural sequence in which the [s] functions as the plural suffix as well as as the onset of the following word (e.g., *kerels soms*) is segmentally identical to a singular sequence in which the [s] is only the onset of the following word (e.g., *kerel soms*): [keʀəlsɔms]. Durational measurements on one speaker's realizations showed that, in the non-ambiguous sentences, the singular form was longer than the stem of the plural form, which is consistent with the findings in Chapters 2 and 3. In the ambiguous sentences, however, the reverse pattern was observed: The singular form was *shorter* than the stem of the plural form. The measurements furthermore showed that the degemination of the cluster of *s*-es was incomplete: The [s] in sequences like *kerels soms* was longer than the [s] in sequences like *kerel soms*.

Listeners' sensitivity to subsegmental cues for morphological complexity

Having established that there are in fact substantial subsegmental differences between a stem in isolation and a stem that is onset-embedded in an inflectional or derivational form, and furthermore between an [s] that underlyingly consists of a single *s* (an onset-*s*) and an [s] that underlyingly consists of a cluster of *s*-es (the plural suffix *-s* and an onset-*s*), this thesis addressed the question of whether listeners are sensitive to these differences.

In Chapter 2, we investigated the combined and the independent effects of durational and intonational information in the speech signal on the processing of singular and plural forms in Dutch [1]. Listeners were presented with forms in which the segmental information mismatched the prosodic information: When the segmental information pointed to the singular form, the prosodic (durational and intonational) information pointed to the plural form, and vice versa. When presented with forms containing mismatching information, listeners' responses were significantly delayed, in a number decision task (where listeners were to decide whether the presented form was in the singular or in the plural) as well as in a lexical decision task. The delay in response times was correlated with the magnitude of the durational mismatch. Furthermore, when only intonational information mismatched the segmental information, the delay in processing was considerably smaller than when both durational and intonational information mismatched the segmental information, but it was still significant. Both durational as well as intonational cues thus appear to be picked up by the listener, and when these sources of information

[1] In the present thesis, we concentrated on the perceptual effects of durational and intonational information in the speech signal. Conceivably, monosyllables differ from the stems of their bisyllabic morphologically related forms in other respects as well, such as in the quality of the vowel and in dynamic spectral information. Subsequent research is needed to uncover the perceptual effects of such differences between monosyllables and the stems of their morphologically related bisyllabic forms.

mismatch the segmental information, number decision as well as lexical decision (for which the number of the noun is irrelevant) are hindered. We argue that in parallel to the processing of the acoustic signal of the stem, an expectation regarding the number of syllables that will follow is built up, based on the relative durations of the segments in the stem. A delay in processing occurs when this expectation is violated by the segmental material that either does or does not follow the stem.

In Chapter 3, we investigated whether we could extend these findings to the processing of stems of comparatives (i.e., inflection) and agent nouns (i.e., derivation), in Dutch as well as in English. It was found that, for agent noun stems as well as for comparative stems, listeners' responses were delayed when the durational information in the stem mismatched the number of syllables of the word that was presented. English is a morphologically poorer language than Dutch. In other words, in English a stem is less often followed by one or more unstressed syllables and thus occurs less often in substantially shortened form. Nevertheless, English listeners are no less sensitive than Dutch listeners to the acoustic cues in the stem that signal whether or not the stem will be followed by one or more unstressed syllables.

Chapter 4 investigated which subsegmental cues — if any — listeners use to resolve the ambiguity between sentences that are segmentally identical, but that differ in the function of the [s]. The [s] functions either only as word-onset (as in *Ik spreek de kerel soms*, with the singular noun *kerel*) or also as the plural suffix of the preceding noun (as in *Ik spreek de kerels soms*, with the plural noun *kerels*). Recall that in the ambiguous sentences, the singular forms were shorter than the stems of the plural forms (which is opposite to what was observed for words in isolation in Chapters 2 and 3, and in the non-ambiguous sentences), and the [s] in sequences like *kerels soms* (in which the [s] functions as the plural suffix of the noun *kerel* as well as as the onset of the following word) was longer than the [s] in sequences like *kerel soms* (in which the [s] functions only as word-onset). In a number decision task, listeners showed performance above chance level for the ambiguous sentences, indicating that they were sensitive to at least some of the acoustic differences between the segmentally identical sentences. Analyses of the response data as well as the reaction time data showed that it was especially the duration of the [s] that listeners used to attempt to resolve the ambiguity between *kerel soms* and *kerels soms*. Listeners also showed sensitivity to the duration of the stem, but this sensitivity led to incorrect responses for the ambiguous sentences, as listeners did not take the reversal of the durational pattern into account: A long stem duration was interpreted as indicating the singular form. A corpus survey showed that the type of ambiguity studied in Chapter 4 occurs very infrequently. Apparently, the durational distributions specific for singulars and stems of plurals in ambiguous sentences are too infrequent for listeners to develop a sensitivity to them.

To conclude, this thesis provides evidence for listeners' capability within the domain of morphology to pick up the fine phonetic details in the acoustic signal. This capability is striking, given the enormous variability in the temporal structure of speech. Chapter 4, however, also shows that there are limits to what patterns of cues the speech-perception system can handle. A corpus survey showed that ambiguous sentences of the type studied here occur very infrequently. Listeners turned out not to be sensitive to the unusual and seldom encountered durational pattern in these sentences: The sensitivity to the duration of the stem had an adverse effect on the comprehension of the ambiguous sentences, as the difference in stem duration between singulars and plurals in the ambiguous sentences went in the opposite direction of the default pattern. We conclude that listeners develop a sensitivity to only those subsegmental patterns that are robustly present in the language.

Listeners' insensitivity to suffix reduction

The question arises how it is possible, given listeners' extraordinary sensitivity to the fine phonetic details in the speech signal, that comprehension does not seem to suffer at all from the drastic reductions that occur in casual speech. In casual speech, words are typically produced with fewer segments, or even syllables, than when they are carefully pronounced in isolation. For example, the Dutch word *eigenlijk*, with the canonical pronunciation [ɛiχələk], may in casual speech be realized as [ɛik] (Ernestus, 2000). When such reduced forms are presented in their natural context, listeners' comprehension is seemingly unaffected by the absence of several segments. When the forms are presented in isolation, however, listeners do not recognize them (Ernestus, Baayen, & Schreuder, 2002).

This phenomenon was studied in Chapter 5. In Experiment 5.1, listeners were presented with speech fragments containing words in which the suffix *-(e)lijk* [(ə)lək] was either completely or partly reduced. In no case was the phoneme [l] realized. In a phoneme-monitoring task, listeners correctly reported the absence of the phoneme [l] when the suffixes — or whatever was left of them — were presented in isolation. When the words containing the reduced suffixes were presented in a context of several words, listeners frequently incorrectly reported the presence of the phoneme [l]. They seemed to 'restore' the suffixes that were either partly or completely missing. Experiment 5.4 showed that the basis for this suffix restoration is mainly phonological in nature, but that there might be a small role for orthography as well. In a phoneme-monitoring task, listeners were presented with stimuli with mismatching orthography and phonology. When explicitly instructed to ignore orthography, listeners still occasionally based their responses on the orthographic representation. The number of false positive responses in this experiment was however significantly smaller than the number of false positive responses in Experiment 5.1, where listeners had to monitor for the phoneme [l] in reduced and non-reduced stimuli. In these stimuli, the phoneme [l] was either present or absent, but orthography and phonology were never in mismatch: If the orthographic representation contained the grapheme *l*, the (canonical) phonological representation contained the corresponding phoneme [l]. The relatively small number of false positive responses in Experiment 5.4 — where restoration could only occur on the basis of orthographic information — as compared to Experiment 5.1, suggests that, although orthography may play a role, phonology is the main source of information that the restoration in Experiment 5.1 was based on.

We conclude that reduced forms activate the canonical representations in the mental lexicon, and that these representations induce reconstruction processes: The activated representations in the lexicon determine what we think we hear. Restoration enables listeners to understand spontaneous speech, without comprehension being hampered at a conscious level by the drastic reductions inherent to this type of speech.

Lexical storage and lexical processing

This thesis presents further evidence that listeners are sensitive to the fine subsegmental details in the speech signal. This raises the question of whether such subsegmental information might be part of the stored lexical representations of words. In Chapter 2, we approached this question by investigating whether the effect of subsegmental mismatch that was observed for existing nouns in a number decision task would be equally strong for pseudowords. We carried out a lexical decision task, and we found that there was indeed an effect of subsegmental mismatch for pseudowords — for which no lexical entries are available —, but that the correlation between the magnitude of the subsegmental mismatch and the delay in processing was stronger for words than for pseudowords. This is consistent with the view that prosodic information is present in the lexical representations of words. We propose that the prosodic mismatch effect for pseudowords reflects the unconditional probabilities for the co-occurrences of segmental durations and syllable structure. In the case of words, these unconditional probabilities might be supplemented by conditional probabilities based on the co-occurrences of the sequence of segments constituting a word's form representation, the durations of these segments, and their syllable structure.

The hypothesis of lexical storage of subsegmental information is also consistent with the pattern of frequency effects observed in the number decision experiments in Chapter 2. This pattern of frequency effects strongly suggest that the subsegmental information in the stem codetermines which of two representations (singular or plural) becomes most active: Prosodic cues for the plural form lead to activation of the plural representation (i.e., to an effect of the plural frequency) whereas prosodic cues for the singular form lead to activation of the singular representation (i.e., to an effect of the singular frequency).

Chapter 3 approached the question of lexical storage of subsegmental information in a different way. In Chapter 3, we studied the effects on processing of two covariates that are word-specific indications of the prevalence of possible continuation forms: Syllable Ratio and Cohort Entropy. Syllable Ratio is the word-specific log odds ratio of observing a phonetically unshortened form versus observing a shortened form. Cohort Entropy is the entropy of the distribution of cohort members at stem-final position. The predictive values of these covariates were evaluated for both Dutch and English. For Dutch, the morphologically richer language, Syllable Ratio emerged as the superior predictor, whereas for English, Cohort Entropy was the better predictor. Apparently, in a language such as English, in which stems occur relatively infrequently in shortened form, listeners are less sensitive to the item-specific distribution of shortened and unshortened stems within the lexicon. Instead, the contents of the cohort codetermine response latencies. The effect of Syllable Ratio that we observed for Dutch is in line with the hypothesis of lexical storage of subsegmental information: We view the effect of Syllable Ratio as an intrinsic part of the process mapping the acoustic input onto the lexicon. The frequency with which the auditory system encounters the inflectional and derivational types over which Syllable Ratio is calculated leaves its traces in the mapping of the acoustic input onto these lexical representations. The hypothesis of lexical storage of prosodically detailed information was also supported by the production data in

Chapter 3, which show that the durational difference between stems in isolation and stems occurring onset-embedded in inflectional or derivational continuation forms is larger for words than for pseudowords.

Chapters 2 and 3 provide data consistent with the view that there is lexical storage of fine phonetic details. These data are therefore in line with exemplar-based theories of lexical processing (e.g., Johnson, 1997; Goldinger, 1998; Bybee, 2001; Pierrehumbert, 2001, 2002, 2003). The data furthermore show that the fine phonetic details in the speech signal reduce the ambiguity between morphologically related words that share their initial morpheme (segments). This finding reduces the competition problem that is the result of having representations for inflected forms in lexical memory (as shown by Baayen, McQueen, Dijkstra, and Schreuder (2003)). In most current models of spoken-word recognition (see, for example, Marslen-Wilson, 1990; Marslen-Wilson, Moss, & Van Halen, 1996; McClelland & Elman, 1986; Norris, 1994), uninflected and inflected forms are cohort competitors. Given the prosodic differences documented in this thesis, the inflected form might well be a less strong cohort competitor for the uninflected form and vice versa.

Contradictorily, listeners show insensitivity to suffix reduction, that is, their comprehension is not hampered when a suffix is either partly or completely reduced. As shown in this thesis, when reduced word forms are heard in their context, the canonical representations of the words become activated, and listeners restore the incomplete or missing suffixes based on these canonical representations. Apparently, not all phonetic variants of words are stored in the mental lexicon, or if they are, they are not equally accessible at all levels. If they were, the phoneme-monitoring responses could have been based on the representation of the reduced form, but in fact they were not. Furthermore, the recognition of reduced word forms in isolation would then not be a problem, but it is (as shown by Ernestus, Baayen, and Schreuder, 2002). It might be argued that the reduced forms are represented in the mental lexicon in contextual collocations (see Sprenger, 2003, for production). This would explain why the forms are recognized when they are presented in their context, but not when they are presented in isolation. The reduced forms under investigation do not typically appear in fixed contexts, however (Ernestus, 2000), rendering the lexical storage of idiomatic contextual collocations unlikely. The activation of the canonical representation upon hearing a reduced form is consistent with models that assume that words have a single representation in the mental lexicon, but these models face the challenge of how to map a highly reduced form such as [ɛik] onto the non-reduced canonical representation [ɛiχələk].

Topics for further research

Several topics addressed in this thesis lend themselves for further investigation.

First, the presence of prosodic differences between stems in isolation and stems onset-embedded in morphological continuation forms should be investigated in a larger and more diverse population of speakers. Evidence for the presence of such differences has been provided in a production study by Baayen et al. (2003), which included four Dutch speakers, and in this thesis, which included two Dutch speakers and one English speaker. Although listeners' sensitivity to these differences in itself provides evidence for the robustness and the consistency of these differences, a large-scale study with speakers from several regional and dialectical backgrounds is necessary in order to reveal whether there might be substantial individual or regional variability in the realization of the prosodic patterns investigated in this thesis.

Interestingly, the two sources of prosodic information that have been shown to influence the processing of singular and plurals in standard Dutch, that is, durational and intonational information, are lexically contrastive in some southern dialects of Dutch. For some nouns in the dialect of Weert, for example, the difference between singular and plural is expressed exclusively by a durational difference (Heijmans & Gussenhoven, 1998), whereas for some nouns in the dialect of Roermond, the difference between singular and plural is expressed exclusively by a tonal contrast (Gussenhoven, 2000). It is conceivable that, as a consequence, speakers of the dialect of Weert would be more sensitive to durational information in the speech signal, whereas speakers of the dialect of Roermond would be more sensitive to intonational information, even in standard Dutch, where durational and intonational information are never lexically contrastive. Future research on how listeners' sensitivity to prosodic information in the speech signal is influenced by their dialectical background is needed.

A third question concerns the time-course of processing prosodic information. Prosodic information can only be truly functional in distinguishing stems from morphologically complex forms if these cues are processed by the perceptual system at an early stage, that is, before segmental information corresponding to the presence or absence of an inflectional or derivational suffix can exert its influence (i.e., can disambiguate the forms). Current research is underway investigating the time-course of the processing of prosodic information in the stem using event-related brain potentials (Kemps, Van Berkum, Ernestus, Schreuder, & Baayen, in preparation).

Fourth, the question remains to what extent listeners' sensitivity to prosodic information in the speech signal plays a role in more natural communicative situations. In this thesis, listeners' sensitivity to prosodic detail has been shown by means of presenting them with 'clean' speech, in rather unnatural tasks (number decision and lexical decision). Future research should elucidate to what extent the prosodic differences reported on in this thesis are actually present in casual speech, and to what extent listeners can make use of such differences.

Finally, it remains an open question how it is possible that highly reduced word forms such as [ɛik] activate their canonical representations ([ɛiχələk]). The fact that they only do so when they are presented in context, but not when they are presented in isolation, suggests a critical role for context. As mentioned above, however, reduced forms do not occur in fixed contexts. It is therefore unlikely that reduced forms are stored in the mental lexicon in contextual collocations. But even if they are, it remains unclear how the canonical form is activated upon hearing a reduced form, and why the activation of the canonical form overrules the activation of the representation of the actual reduced form. This process needs further exploration.

To conclude, this thesis provides evidence for listeners' extraordinary sensitivity within the domain of morphology to the fine phonetic details in the speech signal on the one hand, and listeners' insensitivity to suffix reduction in casual speech on the other hand. Further research should elucidate how to reconcile these contradictory findings.

References

Baayen, R. H., McQueen, J., Dijkstra, T., and Schreuder, R.: 2003, Frequency effects in regular inflectional morphology: Revisiting Dutch plurals, *in* R. H. Baayen and R. Schreuder (eds), *Morphological structure in language processing*, Mouton de Gruyter, Berlin, pp. 355–390.

Bybee, J. L.: 2001, *Phonology and language use*, Cambridge University Press, Cambridge.

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G.: 2002, Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition, *Journal of Experimental Psychology: Human Perception and Performance* **28**, 218–244.

Ernestus, M.: 2000, *Voice assimilation and segment reduction in casual Dutch, a corpus-based study of the phonology-phonetics interface*, LOT, Utrecht.

Ernestus, M., Baayen, R. H., and Schreuder, R.: 2002, The recognition of reduced word forms, *Brain and Language* **81**, 162–173.

Goldinger, S. D.: 1998, Echoes of echoes? An episodic theory of lexical access, *Psychological Review* **105**, 251–279.

Gussenhoven, C.: 2000, The lexical tone contrast of Roermond Dutch in Optimality Theory, *in* M. Horne (ed), *Intonation: Theory and Experiment*, Kluwer, Amsterdam, pp. 129–167.

Heijmans, L. and Gussenhoven, C.: 1998, The Dutch dialect of Weert, *Journal of the International Phonetic Association* **28**, 107–112.

Johnson, K.: 1997, The auditory/perceptual basis for speech segmentation, *Ohio State University Working Papers in Linguistics* **50**, 101–113.

Kemps, R., Van Berkum, J., Ernestus, M., Schreuder, R., and Baayen, R. H.: in preparation, Investigating the time-course of processing acoustic cues to morphological complexity using event-related brain potentials.

Marslen-Wilson, W. D.: 1990, Activation, competition, and frequency in lexical access, *in* G. T. M. Altmann (ed), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, The MIT Press, Cambridge, Mass., pp. 148–172.

Marslen-Wilson, W. D., Moss, H. E., and Van Halen, S.: 1996, Perceptual distance and competition in lexical access, *Journal of Experimental Psychology: Human Perception and Performance* **22**, 1376–1392.

McClelland, J. L. and Elman, J. L.: 1986, The TRACE model of speech perception, *Cognitive Psychology* **18**, 1–86.

Norris, D. G.: 1994, Shortlist: A connectionist model of continuous speech recognition, *Cognition* **52**, 189–234.

Pierrehumbert, J.: 2001, Exemplar dynamics: Word frequency, lenition and contrast, *in* J. Bybee and P. Hopper (eds), *Frequency and the emergence of linguistic structure*, John Benjamins Publishing Company, Amsterdam/Philadelphia, pp. 137–157.

Pierrehumbert, J.: 2002, Word-specific phonetics, *in* C. Gussenhoven and N. Warner (eds), *Laboratory Phonology VII*, Mouton de Gruyter, Berlin, pp. 101–140.

Pierrehumbert, J. B.: 2003, Probabilistic phonology: Discrimination and robustness, *in* R. Bod, J. Hay, and S. Jannedy (eds), *Probability theory in linguistics*, The MIT Press, pp. 177–228.

Salverda, A., Dahan, D., and McQueen, J.: 2003, The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension, *Cognition* **90**, 51–89.

Spinelli, E., McQueen, J., and Cutler, A.: 2003, Processing resyllabified words in French, *Journal of Memory and Language* **48**, 233–254.

Sprenger, S.: 2003, *Fixed expressions and the production of idioms*, MPI Series in Psycholinguistics, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Samenvatting en conclusies

Deze dissertatie beschrijft onderzoek gericht op twee ogenschijnlijk tegenstrijdige kenmerken van het spraakperceptiesysteem. Enerzijds zijn luisteraars uitermate gevoelig voor de subtiele fonetische details in het spraaksignaal. Deze subtiele akoestische cues kunnen de tijdelijke ambiguïteit tussen woorden met initiële segmentele overlap verminderen, en kunnen de lexicale activatie sturen (zie bijvoorbeeld Davis, Marslen-Wilson, & Gaskell, 2002; Spinelli, McQueen, & Cutler, 2003; Salverda, Dahan, & McQueen, 2003). Anderzijds lijkt de herkenning allerminst verstoord te worden door de drastische reducties kenmerkend voor spontane spraak. De herkenning lijkt onaangedaan door het ontbreken van segmenten, en soms zelfs complete syllaben (Ernestus, 2000; Ernestus, Baayen, & Schreuder, 2002). Het primaire doel van het onderzoek beschreven in deze dissertatie was om, door middel van het bestuderen van deze tegenstrijdige verschijnselen in het domein van de morfologie, inzicht te verkrijgen in hoe woorden gerepresenteerd en geactiveerd worden in het mentale lexicon. In Hoofdstuk 2 tot en met 4 werd de rol van subsegmentele cues in de verwerking van morfologische (inflectioneel of derivationeel) gerelateerde woorden met segmentele overlap onderzocht. Hoofdstuk 5 bestudeerde de verwerking van ernstig gereduceerde vormen, in het bijzonder Nederlandse woorden waarin het derivationele suffix *-(e)lijk* [(ə)lək] drastisch gereduceerd is.

Subsegmentele verschillen tussen stammen in isolatie en stammen in morfologisch complexe woorden

Een eerste onderzoeksvraag die in deze dissertatie aan bod kwam is of de subsegmentele verschillen zoals geobserveerd tussen de intiële syllaben van korte woorden en segmenteel overlappende, maar morfologisch *onverwante* langere woorden (bijv., *ham - hamster*) ook consistent optreden tussen de intiële syllaben van korte woorden en segmenteel overlappende, morfologisch *verwante* langere woorden (bijv., *werk - werker*). De morfologische relaties die bestudeerd werden in deze dissertatie zijn regelmatige inflectie en derivatie, in het Nederlands en in het Engels. We vergeleken stammen in isolatie met stammen die gevolgd werden door een inflectioneel of een derivationeel suffix. Het is niet vanzelfsprekend dat de toevoeging van een suffix tot substantiële akoestische verschillen in de stam leidt, aangezien inflectionele en derivationele suffixen zowel in het Engels als in het Nederlands vaak slechts bestaan uit één of twee segmenten (bijv., het Nederlandse meervoudssuffix *-en* wordt gewoonlijk gerealiseerd als slechts een schwa). Hoofdstuk 2, 3, en 4 laten zien dan er desondanks wel degelijk sprake is van substantiële akoestische verschillen in de stam als gevolg van de toevoeging van een inflectioneel of derivationeel suffix.

Hoofdstuk 2 onderzocht of er duur- en intonatieverschillen zijn tussen de enkelvoudsvorm en de stam van de meervoudsvorm van Nederlandse zelfstandige naamwoorden. Monosyllabische zelfstandige naamwoorden die het meervoudssuffix *-en* [ə(n)] nemen werden onderzocht (bijv., *boek - boeken*, [buk] - [bukə(n)]. De realisaties van één spreker van de enkelvoudsvormen en de meervoudsvormen van deze zelfstandige naamwoorden werden bestudeerd. Akoestische metingen aan deze vormen gerealiseerd in isolatie lieten zien dat de stam van de meervoudsvorm (die segmenteel identiek is aan de enkelvoudsvorm) gemiddeld korter was dan de enkelvoudsvorm, en dat de meervoudsvorm een gemiddeld hogere grondfrequentie had dan de enkelvoudsvorm.

In Hoofdstuk 3 werden gelijksoortige duurverschillen geobserveerd tussen monosyllabische adjectieven en hun corresponderende comparatieven (bijv., strikt - strikter), en tussen monosyllabische werkwoordsstammen en de stammen van hun corresponderende agentieven (bijv., werk - werker). Deze duurverschillen traden op in zowel het Nederlands als het Engels (één spreker per taal).

Hoofdstuk 4 onderzocht of overeenkomstige duurverschillen optreden tussen enkelvouden en de stammen van meervouden van Nederlandse zelfstandige naamwoorden die het meervoudssuffix *-s* [s] nemen, ondanks het feit dat toevoeging van dit suffix niet tot toevoeging van een extra syllabe leidt. De enkelvouden en de meervouden werden ingebed in niet-ambigue zinnen (bijv., *Ik spreek de kerel* (enkelvoud)/*kerels* (meervoud) *vaak*) en in ambigue zinnen (bijv., *Ik spreek de kerel/kerels soms*). In de ambigue zinnen werd het zelfstandig naamwoord altijd gevolgd door een woord met een [s]-onset. Als gevolg van degeminatie van het cluster van *s*-en zijn zinnen van dit type met meervouden segmenteel identiek aan zinnen van dit type met enkelvouden. In de meervoudszinnen functioneert de [s] zowel als meervoudssuffix als als onset van het volgende woord (bijv., *kerel soms*), in de enkelvoudszinnen functioneert de [s] alleen als onset van het volgende woord (bijv., *kerel soms*): [keʀəlsɔms]. Duurmetingen aan de realisaties van één spreker lieten zien dat het enkelvoud in de niet-ambigue zinnen langer was dan de stam van het meervoud. Dit is consistent met de bevindingen in Hoofdstuk 2 en 3. In de ambigue zinnen werd echter het tegenovergestelde patroon geobserveerd: De enkelvoudsvorm was *korter* dan de stam van de meervoudsvorm. De metingen lieten verder zien dat de degeminatie van het cluster van *s*-en niet compleet was: De [s] in sequenties als *kerels soms* was langer dan de [s] in sequenties als *kerel soms*.

De gevoeligheid van luisteraars voor subsegmentele cues voor morfologische complexiteit

Na vastgesteld te hebben dat er inderdaad substantiële subsegmentele verschillen zijn tussen een stam in isolatie en een stam die de onset vormt van een inflectionele of derivationele vorm, en bovendien tussen een [s] die onderliggend bestaat uit een enkele *s* (een onset-*s*) en een [s] die onderliggend bestaat uit een cluster van *s*-en (de meervouds-*s* en een onset-*s*), behandelde deze dissertatie de vraag of luisteraars gevoelig zijn voor deze verschillen.

In Hoofdstuk 2 onderzochten we de gecombineerde en de onafhankelijke effecten van duur- en intonatie-informatie in het spraaksignaal op de verwerking van enkelvouden en meervouden in het Nederlands [2]. Wij presenteerden luisteraars met vormen waarin de segmentele informatie niet overeenkwam met de prosodische informatie: Wanneer de segmentele informatie op het enkelvoud wees, wees de prosodische (durationele en intonationele) informatie op het meervoud, en vice versa. De responsen van luisteraars op dergelijke vormen waren significant vertraagd ten opzichte van hun responsen op normale vormen, in zowel een getalsbeslissingstaak (waarin luisteraars geacht werden te beslissen of de gepresen-

[2]In deze dissertatie concentreerden wij ons op de perceptuele effecten van duur- en intonatie-informatie in het spraaksignaal. Mogelijk verschillen monosyllaben ook in andere opzichten van de stammen van hun bisyllabische morfologisch gerelateerde vormen, zoals in klinkerkwaliteit of in dynamisch spectrale informatie. Nader onderzoek moet uitwijzen wat de perceptuele effecten van dergelijke verschillen tussen monosyllaben en de stammen van hun bisyllabische morfologisch gerelateerde vormen zijn.

teerde vorm een enkelvoud of een meervoud was) als in lexicale decisie. De hoeveelheid vertraging was gecorreleerd met de grootte van de durationele mismatch. Wanneer alleen de intonationele informatie niet overeenkwam met de segmentele informatie, was de vertraging in de verwerking substantieel kleiner dan wanneer zowel durationele als intonationele informatie niet overeenkwamen met de segmentele informatie, maar de vertraging was nog steeds significant. Zowel de durationele als de intonationele cues lijken dus benut te worden door de luisteraar, en wanneer deze bronnen van informatie niet overeenkomen met de segmentele informatie, worden zowel getalsbeslissing als lexicale decisie (waarvoor het getal van het zelfstandig naamwoord irrelevant is) gehinderd. We argumenteren dat er, parallel aan de verwerking van het akoestische signaal corresponderend met de stam, een verwachting wordt opgebouwd over het aantal volgende syllaben, gebaseerd op de relatieve duren van de segmenten in de stam. Wanneer aan deze verwachting niet voldaan wordt door het segmentele materiaal dat al dan niet volgt op de stam, treedt er een vertraging in de verwerking op.

In Hoofdstuk 3 onderzochten we of we deze bevindingen konden verbreden naar de verwerking van stammen van comparatieven (i.e., inflectie) en agentieven (i.e., derivatie), in zowel het Nederlands als het Engels. Voor stammen van zowel comparatieven als agentieven vonden we dat de responsen van luisteraars vertraagd werden wanneer de durationele informatie in de stam niet in overeenstemming was met het aantal syllaben van het gepresenteerde woord. Engels is een morfologisch minder rijke taal dan Nederlands. Met andere woorden, een stam wordt minder vaak gevolgd door één of meer onbeklemtoonde syllaben, en komt dus minder vaak voor in substantieel verkorte vorm in het Engels dan in het Nederlands. Desondanks bleken Engelse luisteraars niet minder gevoelig dan Nederlandse luisteraars voor de akoestische cues in de stam die aangeven of de stam al dan niet gevolgd wordt door één of meer onbeklemtoonde syllaben.

Hoofdstuk 4 onderzocht of en — zo ja — welke subsegmentele cues luisteraars gebruiken voor het oplossen van de ambiguïteit tussen zinnen die segmenteel identiek zijn, maar die verschillen wat betreft de functie van de [s]. De [s] functioneert ofwel alleen als woordonset (zoals in *Ik spreek de kerel soms*, met de enkelvoudsvorm *kerel*) ofwel ook als het meervoudssuffix van het voorafgaande zelfstandige naamwoord (zoals in *Ik spreek de kerels soms*, met de meervoudsvorm *kerels*). In de ambigue zinnen bleken de enkelvouden korter dan de stammen van de meervouden (hetgeen tegengesteld is aan het patroon geobserveerd voor woorden in isolatie in Hoofdstuk 2 en 3, en voor de niet-ambigue zinnen), en de [s] in sequenties zoals *kerels soms* (waarin de [s] zowel als meervoudssuffix van het zelfstandige naamwoord als als onset van het volgende woord functioneert) bleek langer dan de [s] in sequenties zoals *kerel soms* (waarin de [s] alleen als woord-onset functioneert). In een getalsbeslissingstaak presteerden luisteraars boven kansnivo voor de ambigue zinnen, hetgeen aangeeft dat de luisteraars gevoelig waren voor tenminste enkele akoestische verschillen tussen de segmenteel identieke zinnen. Analyses van zowel de responsdata als de reactietijden toonden aan dat met name de duur van de [s] door de luisteraars gebruikt werd in een poging de ambiguïteit tussen *kerel soms* en *kerels soms* op te lossen. De luisteraars toonden eveneens gevoeligheid voor de duur van de stam, maar deze gevoeligheid leidde tot incorrecte responsen voor de ambigue zinnen, aangezien luisteraars de omkering van het durationele patroon niet in aanmerking namen: Een lange stamduur werd geïnterpreteerd als evidentie voor de enkelvoudsvorm. Een corpusstudie wees uit dat het type ambiguïteit zoals onderzocht in Hoofdstuk 4 zeer zelden voorkomt. Blijkbaar zijn de durationele distributies specifiek voor enkelvouden en de stammen van meervouden in ambigue zinnen te zeldzaam om tot gevoeligheid van de luisteraar voor deze distributies te leiden.

Samenvattend, deze dissertatie beschrijft evidentie binnen het domein van de morfologie voor het vermogen van luisteraars om de subtiele fonetische details in het akoestische signaal te benutten. Dit vermogen is frappant, gegeven het feit dat de temporele structuur van spraak uitermate variabel is. Hoofdstuk 4 laat echter zien dat er ook grenzen zijn aan welke patronen van cues het spraakperceptiesysteem kan benutten. Een corpusstudie toonde aan dat het type ambigue zinnen bestudeerd in dit hoofdstuk zeer zelden optreedt. Luisteraars bleken niet gevoelig voor het ongebruikelijke en zeldzame durationele patroon in deze zinnen: De gevoeligheid voor de duur van de stam had een negatief effect op de herkenning van de ambigue zinnen, omdat de richting van het verschil in stamduur tussen enkelvouden en meervouden in de ambigue zinnen tegengesteld was aan de gebruikelijke richting. We concluderen dat luisteraars gevoeligheid ontwikkelen voor alleen die subsegmentele patronen die robuust aanwezig zijn in de taal.

De ongevoeligheid van luisteraars voor reductie van het suffix

De vraag doet zich voor hoe het mogelijk is, gegeven de buitengewone gevoeligheid van luisteraars voor de subtiele akoestische details in het spraaksignaal, dat de herkenning allerminst gehinderd lijkt te worden door de dratische reducties die optreden in spontane spraak. In spontane spraak worden woorden gewoonlijk geproduceerd met minder segmenten, of soms zelfs minder syllaben, dan wanneer de ze nauwkeurig worden uitgesproken in isolatie. Het Nederlandse woord *eigenlijk*, bijvoorbeeld, met de kanonieke uitspraak [ɛiχələk], kan in spontane spraak gerealiseerd worden als [ɛik] (Ernestus, 2000). Wanneer dergelijke gereduceerde vormen in hun natuurlijke context gepresenteerd worden, lijkt de herkenning onaangedaan door het ontbreken van enkele segmenten. Wanneer de vormen echter in isolatie worden gepresenteerd, worden ze niet herkend (Ernestus, Baayen, & Schreuder, 2002).

Dit fenomeen werd bestudeerd in Hoofdstuk 5. In Experiment 5.1 boden wij luisteraars spraakfragmenten aan die woorden bevatten waarin het suffix *-(e)lijk* [(ə)lək] geheel dan wel gedeeltelijk gereduceerd was. Het foneem [l] was in geen geval gerealiseerd. In een foneemdetectietaak rapporteerden luisteraars terecht de afwezigheid van het foneem [l] wanneer de suffixen — of datgene wat ervan resteerde — werden aangeboden in isolatie. Wanneer de woorden met de gereduceerde suffixen echter werden aangeboden in een context van enkele woorden, rapporteerden de luisteraars vaak ten onrechte de aanwezigheid van het foneem [l]. De luisteraars leken de geheel of gedeeltelijk afwezige suffixen te ‘restaureren’. Experiment 5.4 liet zien dat deze suffixrestauratie met name gebaseerd is op fonologische informatie, maar dat orthografische informatie eveneens een kleine rol zou kunnen spelen. In een foneemdetectietaak werden stimuli aangeboden waarvan de orthografie niet in overeenstemming was met de fonologie. Wanneer de luisteraars expliciet werden geïnstrueerd de orthografie te negeren, baseerden zij desondanks enkele responsen op de orthografische representaties. Het aantal ‘false positives’ in dit experiment was echter significant kleiner dan het aantal ‘false positives’ in Experiment 5.1, waarin luisteraars de aanwezigheid dan wel afwezigheid van het foneem [l] moesten detecteren in gereduceerde en ongereduceerde stimuli. In deze stimuli was het foneem [l] aanwezig dan wel afwezig, maar orthografie en fonologie waren altijd met elkaar in overeenstemming: Als de orthografische representatie het grafeem *l* bevatte, bevatte de (kanonieke) fonologische representatie het corresponderende foneem [l]. Het relatief kleine aantal 'false positives' in Experiment 5.4 — waar restauratie alleen op basis van orthografie kon plaatsvinden — vergeleken met Experiment 5.1 suggereert dat, ook al zou orthografie een rol kunnen spelen, fonologie de primaire bron van informatie is waarop de restauratie in Experiment 5.1 gebaseerd werd.

We concluderen dat gereduceerde vormen de kanonieke representaties in het mentale lexicon activeren, en dat deze representaties reconstructieprocessen induceren: De geactiveerde representaties in het lexicon bepalen wat we denken te horen. Restauratie stelt luisteraars in staat om spontane spraak te begrijpen zonder dat de herkenning op een bewust nivo gehinderd wordt door de drastische reducties die inherent zijn aan dit type spraak.

Lexicale opslag en lexicale verwerking

Deze dissertatie beschrijft additionele evidentie dat luisteraars gevoelig zijn voor de subtiele subsegmentele details in het spraaksignaal. Dit roept de vraag op of dergelijke subsegmentele informatie wellicht onderdeel uitmaakt van de opgeslagen lexicale representaties van woorden. In Hoofdstuk 2 benaderden we deze vraag middels het toetsen of het subsegmentele mismatch effect zoals geobserveerd voor bestaande zelfstandige naamwoorden in een getalsbeslissingstaak even sterk optreedt voor pseudowoorden. We ondernamen een lexicaal decisie-experiment en we vonden dat er inderdaad een effect van subsegmentele mismatch optreedt voor pseudowoorden — die geen lexicale representatie hebben— maar dat de correlatie tussen de grootte van de subsegmentele mismatch en de vertraging in de verwerking sterker was voor woorden dan voor pseudowoorden. Deze bevinding is consistent met de visie dat prosodische informatie onderdeel uitmaakt van de lexicale representaties van woorden. We stellen voor dat het prosodisch mismatch effect voor pseudowoorden een reflectie is van de onvoorwaardelijke waarschijnlijkheden van het samen optreden van segmentduren en syllabische structuur. In het geval van woorden worden deze onvoorwaardelijke waarschijnlijkheden mogelijk aangevuld met voorwaardelijke waarschijnlijkheden gebaseerd op het samen optreden van de sequentie van segmenten die de vormrepresentatie van een woord vormen, de duren van deze segmenten en hun syllabische structuur.

De hypothese dat subsegmentele informatie lexicaal is opgeslagen is ook consistent met het patroon van frequentie-effecten zoals geobserveerd in de getalsbeslissingsexperimenten in Hoofdstuk 2. Dit patroon van frequentie-effecten suggereert dat de subsegmentele informatie in de stam mede bepaalt welke van twee representaties (de enkelvouds- of meervoudsrepresentatie) het meest geactiveerd wordt: Prosodische cues voor de meervoudsvorm leiden tot activatie van de meervoudsrepresentatie (i.e., tot een effect van de meervoudsfrequentie) terwijl prosodische cues voor de enkelvoudsvorm leiden tot activatie van de enkelvoudsrepresentatie (i.e., tot een effect van de enkelvoudsfrequentie).

Hoofdstuk 3 benaderde de kwestie van lexicale opslag van subsegmentele informatie op een andere manier. In Hoofdstuk 3 bestudeerden we de effecten op de verwerking van twee covariaten die een woord-specifieke indicatie geven van het voorkomen van mogelijke continueringsvormen: Syllable Ratio en Cohort Entropy. Syllable Ratio is de woord-specifieke log-odds ratio van het observeren van een fonetisch onverkorte vorm versus het observeren van een verkorte vorm. Cohort Entropy is de entropie van de verdeling van de leden van het cohort op stamfinale positie. De predictieve waarden van deze covariaten werden geëvalueerd voor zowel het Nederlands als het Engels. Voor het Nederlands, de morfologisch rijkere taal, bleek Syllable Ratio de betere voorspeller, terwijl Cohort Entropy voor het Engels de betere voorspeller was. Blijkbaar zijn luisteraars in een taal zoals het Engels, waarin stammen relatief zelden voorkomen in verkorte vorm, minder gevoelig voor de item-specifieke verdeling van verkorte en onverkorte stammen in het lexicon. In plaats daarvan bepaalt de inhoud van het cohort mede de reactietijden. Het effect van Syllable Ratio zoals geobserveerd voor het Nederlands is consistent met de hypothese dat subsegmentele informatie lexicaal is opgeslagen: We beschouwen het effect van Syllable Ratio als een intrinsiek onderdeel van het proces dat de akoestische input afbeeldt op het lexicon. De frequentie waarmee het auditieve systeem in aanraking komt met de inflectionele en derivationele types waarover Syllable Ratio berekend wordt, laat sporen na in de afbeelding van de akoestische input op deze lexicale representaties. De hypothese dat prosodisch gedetailleerde informatie lexicaal opgeslagen is, werd eveneens ondersteund door de productiedata in Hoofdstuk 3, die laten zien dat het duurverschil tussen stammen in isolatie en stammen die de onset vormen van inflectionele en derivationele continueringsvormen groter is voor woorden dan voor pseudowoorden.

Hoofdstuk 2 en 3 rapporteren data consistent met de visie dat subtiele fonetische details lexicaal zijn opgeslagen. Deze data zijn daarom coherent met ‘exemplar’-gebaseerde theorieën met betrekking tot lexicale verwerking (bijv., Johnson, 1997; Goldinger, 1998; Bybee, 2001; Pierrehumbert, 2001, 2002, 2003). De data laten bovendien zien dat de subtiele fonetische details in het spraaksignaal de ambiguïteit reduceren tussen morfologisch gerelateerde woorden die het initiële morfeem (segmenten) delen. Deze bevinding reduceert het competitieprobleem dat het resultaat is van de aanwezigheid van representaties voor geïnflecteerde vormen in het lexicale geheugen (zoals aangetoond door Baayen, McQueen, Dijkstra, en Schreuder (2003)). In de meeste gangbare modellen van gesproken woordherkenning (zie bijvoorbeeld Marslen-Wilson, 1990; Marslen-Wilson, Moss, & Van Halen, 1996; McClelland & Elman, 1986; Norris, 1994) zijn ongeïnflecteerde en geïnflecteerde vormen cohort-concurrenten. Dankzij de prosodische verschillen gerapporteerd in deze dissertatie, zijn de geïnflecteerde en de ongeïnflecteerde vorm wellicht minder sterke concurrenten.

Ogenschijnlijk tegenstrijdig met hun gevoeligheid voor de details in het spraaksignaal, vertonen luisteraars geen gevoeligheid voor reductie van het suffix: De herkenning is allerminst verstoord wanneer een suffix geheel of gedeeltelijk gereduceerd is. Deze dissertatie laat zien dat, wanneer gereduceerde vormen gehoord worden in hun context, de kanonieke representaties van de woorden geactiveerd worden en dat luisteraars de incomplete of ontbrekende suffixen restaureren op de basis van deze kanonieke representaties. Blijkbaar zijn ofwel niet alle fonetische varianten van woorden opgeslagen in het mentale lexicon, ofwel — als ze wel zijn opgeslagen — zijn niet alle fonetische varianten gelijkelijk toegankelijk op alle nivo's van de lexicale verwerking. Als ze dat wel waren, dan hadden de foneemdetectieresponsen gebaseerd kunnen worden op de representaties van de gereduceerde vormen. Bovendien zou de herkenning van gereduceerde vormen in isolatie dan geen probleem mogen vormen. Dit bleek echter wel het geval (zoals aangetoond door Ernestus, Baayen, en Schreuder, 2002). Men zou kunnen argumenteren dat de gereduceerde vormen gerepresenteerd zijn in het mentale lexicon als contextuele collocaties (zie Sprenger, 2003, voor productie). Dit zou verklaren waarom de vormen herkend worden wanneer ze in hun context worden aangeboden, maar niet wanneer ze in isolatie worden aangeboden. De gereduceerde vormen bestudeerd in deze dissertatie komen echter niet typisch voor in vaste contexten (Ernestus, 2000), hetgeen de lexicale opslag van idiomatische contextuele collocaties onwaarschijnlijk maakt. De activatie van de kanonieke representatie bij het horen van een gereduceerde vorm is consistent met modellen die veronderstellen dat woorden slechts één representatie in het mentale lexicon hebben, maar het is onduidelijk hoe dergelijke modellen een ernstig gereduceerde vorm zoals [ɛik] afbeelden op de ongereduceerde kanonieke representatie [ɛiχələk].

Ter conclusie, deze dissertatie beschrijft evidentie voor de buitengewone gevoeligheid van luisteraars voor de subtiele fonetische details in het spraaksignaal binnen het domein van de morfologie enerzijds, en voor de ongevoeligheid van luisteraars voor reductie van het suffix in spontane spraak anderzijds. Nader onderzoek moet uitwijzen hoe deze tegenstrijdige bevindingen verzoend kunnen worden.

References

Baayen, R. H., McQueen, J., Dijkstra, T., and Schreuder, R.: 2003, Frequency effects in regular inflectional morphology: Revisiting Dutch plurals, *in* R. H. Baayen and R. Schreuder (eds), *Morphological structure in language processing*, Mouton de Gruyter, Berlin, pp. 355–390.

Bybee, J. L.: 2001, *Phonology and language use*, Cambridge University Press, Cambridge.

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G.: 2002, Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition, *Journal of Experimental Psychology: Human Perception and Performance* **28**, 218–244.

Ernestus, M.: 2000, *Voice assimilation and segment reduction in casual Dutch, a corpus-based study of the phonology-phonetics interface*, LOT, Utrecht.

Ernestus, M., Baayen, R. H., and Schreuder, R.: 2002, The recognition of reduced word forms, *Brain and Language* **81**, 162–173.

Goldinger, S. D.: 1998, Echoes of echoes? An episodic theory of lexical access, *Psychological Review* **105**, 251–279.

Johnson, K.: 1997, The auditory/perceptual basis for speech segmentation, *Ohio State University Working Papers in Linguistics* **50**, 101–113.

Marslen-Wilson, W. D.: 1990, Activation, competition, and frequency in lexical access, *in* G. T. M. Altmann (ed), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, The MIT Press, Cambridge, Mass., pp. 148–172.

Marslen-Wilson, W. D., Moss, H. E., and Van Halen, S.: 1996, Perceptual distance and competition in lexical access, *Journal of Experimental Psychology: Human Perception and Performance* **22**, 1376–1392.

McClelland, J. L. and Elman, J. L.: 1986, The TRACE model of speech perception, *Cognitive Psychology* **18**, 1–86.

Norris, D. G.: 1994, Shortlist: A connectionist model of continuous speech recognition, *Cognition* **52**, 189–234.

Pierrehumbert, J.: 2001, Exemplar dynamics: Word frequency, lenition and contrast, *in* J. Bybee and P. Hopper (eds), *Frequency and the emergence of linguistic structure*, John Benjamins Publishing Company, Amsterdam/Philadelphia, pp. 137–157.

Pierrehumbert, J.: 2002, Word-specific phonetics, *in* C. Gussenhoven and N. Warner (eds), *Laboratory Phonology VII*, Mouton de Gruyter, Berlin, pp. 101–140.

Pierrehumbert, J. B.: 2003, Probabilistic phonology: Discrimination and robustness, *in* R. Bod, J. Hay, and S. Jannedy (eds), *Probability theory in linguistics*, The MIT Press, pp. 177–228.

Salverda, A., Dahan, D., and McQueen, J.: 2003, The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension, *Cognition* **90**, 51–89.

Spinelli, E., McQueen, J., and Cutler, A.: 2003, Processing resyllabified words in French, *Journal of Memory and Language* **48**, 233–254.

Sprenger, S.: 2003, *Fixed expressions and the production of idioms*, MPI Series in Psycholinguistics, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Curriculum Vitae

Rachèl Kemps was born in Heerlen, The Netherlands, on August 23, 1976. She studied Clinical Neuropsychology and Speech and Language Pathology at the University of Nijmegen, The Netherlands, and received her respective M.A. degrees in 2000 and 2002. In 2000, she worked as a research assistent in the Comprehension Group at the Max Planck Institute for Psycholinguistics as well as in in the research project 'Early precursors of familial dyslexia: A prospective longitudinal study', funded by the Dutch Research Council (NWO). In January 2001, she began as a Ph.D student in the PIONIER project 'The balance of storage and computation in the mental lexicon', funded by the Dutch Research Council (NWO), the Max Planck Institute for Psycholinguistics, and the Faculty of Arts of the University of Nijmegen. Since January 2004, she works as a postdoctoral research fellow in the Centre for Comparative Psycholinguistics — Department of Linguistics at the University of Alberta, Canada.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turennout*

2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels Schiller*

3. Lexical access in the production of ellipsis and pronouns. *Bernadette Schmitt*

4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*

5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*

6. Gesture and speech production. *Jan-Peter de Ruiter*

7. Comparative intonational phonology: English and German. *Esther Grabe*

8. Finiteness in adult and child German. *Ingeborg Lasser*

9. Language input for word discovery. *Joost van de Weijer*

10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*

11. Producing past and plural inflections. *Dirk Janssen*

12. Valence and transitivity in Saliba, an Austronesian language of Papua New Guinea. *Anna Margetts*

13. From speech to words. *Arie van der Lugt*

14. Simple and complex verbs in Jaminjung: A study of event categorization in an Australian language. *Eva Schultze-Berndt*

15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*

16. Language-specific listening: The case of phonetic sequences. *Andrea Weber*

17. Moving eyes and naming objects. *Femke van der Meulen*

18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*

19. Morphology in speech comprehension. *Kerstin Mauth*

20. Morphological families in the mental lexicon. *Nivja H. de Jong*

21. Fixed expressions and the production of idioms. *Simone A. Sprenger*

22. The grammatical encoding of postural semantics in Goemai (a West Chandic language of Nigeria). *Birgit Hellwig*

23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*

24. Contextual influences on spoken-word processing: An electrophysiological approach. *Daniëlle van den Brink*

25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*

26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*

27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*

28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*